ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

ON WARING'S PROBLEM FOR FOURTH POWERS

BY H. DAVENPORT

(Received April 21, 1939)

Introduction

The object of this paper is to give proofs of the following theorems:

Every sufficiently large number is representable as the sum of 14 *fourth powers, unless it is congruent to* 15 *or* 16 (*mod* 16).

Every sufficiently large number is representable as the sum of 15 *fourth powers, unless it is of the form* $16^h k$, *where k has one of a finite number of values.*

Every sufficiently large number is representable as the sum of 16 *fourth powers.*

The second and third of these theorems are immediate consequences of the first. It is well known that the third theorem is the best possible result of its kind, since $16^h \times 31$ is not the sum of less than 16 fourth powers.

These results are a considerable improvement on the theorem (proved simultaneously by Estermann and by Heilbronn and myself) that every large number is the sum of 17 fourth powers. The improvement results from the use of a new method, of which I have given accounts elsewhere,[1] for constructing different numbers which are the sums of s positive integral k^{th} powers. The reader who is familiar with Waring's problem will see from the proof of Lemma 1 exactly what the new idea is, and will probably be content to take the rest of the paper for granted. But for the sake of other readers, the whole proof has been presented in fairly complete detail.

The only external sources referred to are: (1) Landau, *Vorlesungen über Zahlentheorie* (Leipzig, 1927), volume 1; (2) Davenport and Heilbronn, "On an exponential sum," *Proc. London Math. Soc.*, 41 (1936), 449–453.

The fundamental lemmas

Throughout the paper, all small Latin letters (with or without suffixes) except e, i, f, denote positive integers, and ϵ denotes an arbitrarily small positive number. In the present section, the constants implied by the symbol O depend only on ϵ.

LEMMA 1.[2] *Suppose that P is a large positive integer, and that $u_1 < u_2 < \cdots < u_U < P^{\mu+3}$, where $0 < \mu \leqq \frac{1}{2}$. Suppose also that*

$$U > P^{3(1-\mu)-\epsilon}. \tag{1}$$

[1] *Proc. Royal Soc.*; *Acta Arithmetica.* An account of an alternative method, due to Erdös, is in course of publication in these *Annals*. This method yields a result for four fourth powers which would allow one to prove the third of the theorems enunciated above, but not the first two.

[2] The general form of the result will be found in the author's paper "On sums of positive integral k^{th} powers," *Acta Arithmetica.*

Then the number of solutions of

$$x^4 + u_h = y^4 + u_j, \tag{2}$$

subject to

$$P \leqq x \leqq 2P, \qquad P \leqq y \leqq 2P, \tag{3}$$

is

$$O(P^2U^2P^{3\mu-4+2\epsilon}).$$

PROOF. The number of solutions with $x = y$ is obviously

$$O(PU) = O(P^2U^2P^{3\mu-4+\epsilon}),$$

by (1). Hence it suffices to consider solutions with $y > x$. Writing $y = x + t$, (2) becomes

$$4x^3t + 6x^2t^2 + 4xt^3 + t^4 + u_j = u_h. \tag{4}$$

Plainly

$$P^3t \leqq x^3t < u_h < P^{\mu+3},$$

whence

$$t < P^\mu. \tag{5}$$

Denote by M_1 the number of solutions of (4), subject to (3), (5), and by $M(t, u_h)$ the number of solutions for given values of t, u_h. By Cauchy's inequality,

$$M_1 = \sum_{t,u_h} M(t, u_h) \leqq \Big(\sum_{t,u_h} 1\Big)^{\frac{1}{2}} \Big(\sum_{t,u_h} M^2(t, u_h)\Big)^{\frac{1}{2}}$$

$$< (P^\mu U)^{\frac{1}{2}}(M_1')^{\frac{1}{2}},$$

where M_1' denotes the number of solutions of

$$4x^3t + \cdots + t^4 + u_j = 4x'^3t + \cdots + t^4 + u_{j'} = u_h, \tag{6}$$

in which x' is subject to the same inequality as x. The number of solutions with $x' = x$ is precisely M_1. In the solutions with $x' > x$, we write $x' = x + t_1$. (6) implies

$$tt_1(12x^2 + 12x(t + t_1) + 4t^2 + 6tt_1 + 4t_1^2) + u_{j'} = u_j, \tag{7}$$

and t_1 is subject to $t_1 \leqq P$. Hence

$$M_1' \leqq M_1 + 2M_2 = O(\max(M_1, M_2)),$$

where M_2 denotes the number of solutions of (7). Thus

$$M_1 = O(P^\mu U) + O(P^\mu UM_2)^{\frac{1}{2}}. \tag{8}$$

We now repeat on (7) the argument which was applied to (4). Let $M(t, t_1, u_j)$ denote the number of solutions of (7) for given values of t, t_1, u_j. By Cauchy's inequality,

$$M_2 = \sum_{t,t_1,u_j} M(t, t_1, u_j) \leqq \Big(\sum_{t,t_1,u_j} 1\Big)^{\frac{1}{2}} \Big(\sum_{t,t_1,u_j} M^2(t, t_1, u_j)\Big)^{\frac{1}{2}}$$
$$< (P^{\mu+1}U)^{\frac{1}{2}}(M_2')^{\frac{1}{2}},$$

where M_2' denotes the number of solutions of

(9) $$tt_1(12x^2 + \cdots) + u_{j'} = tt_1(12x'^2 + \cdots) + u_{j''} = u_j .$$

The number of solutions with $x' = x$ is precisely M_2. In the solutions with $x' > x$ we write $x' = x + t_2$. (9) implies

(10) $$12tt_1t_2(2x + t + t_1 + t_2) = u_{j'} - u_{j''},$$

and t_2 is subject to $t_2 \leqq P$. Hence

$$M_2' \leqq M_2 + 2M_3 = O\,(\max\,(M_2, M_3)),$$

where M_3 denotes the number of solutions of (10). Thus

(11) $$M_2 = O(P^{\mu+1}U) + O(P^{\mu+1}UM_3)^{\frac{1}{2}}.$$

Given $u_{j'}, u_{j''}$, the equation (10) allows only $O(P^{\epsilon})$ possibilities for t, t_1, t_2, $2x + t + t_1 + t_2$ as factors of $u_{j'} - u_{j''}$.[3] Hence

(12) $$M_3 = O(P^{\epsilon}U^2).$$

By (8), (11), (12),

$$M_1 = O(P^{\mu}U) + O(P^{\frac{1}{2}\mu}U^{\frac{1}{2}}\{P^{\mu+1}U + P^{\frac{1}{2}\mu+\frac{1}{2}+\epsilon}U^{3/2}\}^{\frac{1}{2}})$$
$$= O(P^{\mu+\frac{1}{2}}U) + O(P^{\frac{3}{4}\mu+\frac{1}{4}+\epsilon}U^{5/4}).$$

Since $\mu \leqq \frac{1}{2}$,

$$P^{\mu+\frac{1}{2}}U \leqq PU = O(P^2U^2P^{3\mu-4+\epsilon}).$$

Also, by (1),

$$P^{\frac{3}{4}\mu+\frac{1}{4}+\epsilon}U^{5/4} = O(P^{\frac{3}{4}\mu+\frac{1}{4}+\epsilon}U^2P^{-(9/4)(1-\mu)+\frac{3}{4}\epsilon})$$
$$= O(P^2U^2P^{3\mu-4+2\epsilon}).$$

This proves Lemma 1.

LEMMA 2. *Suppose that* $s \geqq 2$, *and that* f *is one of* $0, 1, 2, \cdots, s$. *For* $n > n_0(\epsilon)$, *there exist at least* $n^{\gamma_s-\epsilon}$ *numbers less than* n *representable as the sum of* s *fourth powers and congruent to* f (mod (16), *where*

(13) $$\gamma_2 = \frac{1}{2}, \qquad \gamma_3 = \frac{19}{28}, \qquad \gamma_4 = \frac{331}{412}, \cdots, \qquad \gamma_s = \frac{3 + 13\gamma_{s-1}}{4(3 + \gamma_{s-1})}.$$

[3] Landau, Satz 261.

PROOF. (1) $s = 2$. Let $f = f_1 + f_2$, where $f_1 = 0$ or 1 and $f_2 = 0$ or 1. The number of pairs x_1, x_2 satisfying

$$x_1 < (\tfrac{1}{2}n)^{\frac{1}{4}}, \qquad x_2 < (\tfrac{1}{2}n)^{\frac{1}{4}}, \qquad x_1 \equiv f_1 \pmod 2, \qquad x_2 \equiv f_2 \pmod 2$$

is greater than $Cn^{\frac{1}{2}}$, where C is a positive absolute constant. It is well known[4] that the number of representations of an integer m as $x_1^4 + x_2^4$ is $O(m^{\epsilon})$. Hence there are at least

$$\frac{Cn^{\frac{1}{2}}}{O(n^{\epsilon})} > n^{\frac{1}{2}-2\epsilon}$$

numbers less than n representable as the sum of 2 fourth powers and $\equiv f_1^4 + f_2^4 \equiv f \pmod{16}$.

(2) $s > 2$. We assume that the assertion of the lemma is true for $s - 1$, with $\frac{1}{2} \leqq \gamma_{s-1} < 1$, and deduce that it is true for s where γ_s is given by the last formula of (13).

Let $f = f_1 + f_2$, where f_1 is one of $0, 1, \cdots, s - 1$, and f_2 is 0 or 1. Let

$$\mu = \frac{3(1 - \gamma_{s-1})}{3 + \gamma_{s-1}}. \tag{14}$$

Since $\frac{1}{2} \leqq \gamma_{s-1} < 1$, we have $0 < \mu < \frac{1}{7}$. Let $P = [\frac{1}{2}(\frac{1}{2}n)^{\frac{1}{4}}]$. Let $u_1 < u_2 < \cdots < u_U < P^{\mu+3}$ be the numbers less than $P^{\mu+3}$ representable as the sum of $s - 1$ fourth powers and congruent to $f_1 \pmod{16}$. By hypothesis

$$\begin{aligned} U &> P^{(\mu+3)\gamma_{s-1}-\epsilon} \\ &= P^{3(1-\mu)-\epsilon}, \end{aligned}$$

by (14).

Let $r(m)$ denote the number of representations of m as $x^4 + u_h$, where $P \leqq x \leqq 2P$, and $x \equiv f_2 \pmod 2$. Plainly

$$\sum_m r(m) \geqq \tfrac{1}{2}PU.$$

Also $\sum_m r^2(m)$ does not exceed the number of solutions of (2) subject to (3). Since the conditions of Lemma 1 are satisfied,

$$\sum_m r^2(m) = O(P^2 U^2 P^{3\mu-4+2\epsilon}).$$

The number of numbers less than n representable as the sum of s fourth powers and congruent to $f \pmod{16}$ is

$$\begin{aligned} \geqq \sum_{\substack{m \\ r(m)>0}} 1 \geqq \frac{(\sum_m r(m))^2}{\sum_m r^2(m)} &> P^{4-3\mu-3\epsilon} \\ &> n^{\gamma_s - \epsilon}, \end{aligned}$$

[4] Landau, Satz 262.

where

$$\gamma_s = \tfrac{1}{4}(4 - 3\mu) = \frac{3 + 13\gamma_{s-1}}{4(3 + \gamma_{s-1})},$$

by (14).

Notation

Let N be the large integer, not congruent to 15 or 16 (mod 16), which is to be represented as the sum of 14 fourth powers.

There exists an integer $f = 0, 1, 2, 3,$ or 4, such that

(15) $$N - 2f \equiv 1, 2, 3, 4, 5, \text{ or } 6 \pmod{16}.$$

Let

(16) $$P = \left[\left(\frac{N}{100}\right)^{\frac{1}{4}}\right],$$

and let

(17) $$\mu = \frac{243}{1567}.$$

Let $u_1 < u_2 < \cdots < u_U < P^{\mu+3}$ be the numbers less than $P^{\mu+3}$ representable as the sum of 4 fourth powers and congruent to f (mod 16). By Lemma 2,

(18) $$U > P^{(\mu+3)\gamma_4 - \epsilon}, \qquad \gamma_4 = \frac{331}{412}.$$

We observe that

(19) $$(\mu + 3)\gamma_4 = 3(1 - \mu).$$

For any real α, let

$$T(\alpha) = \sum_{x=P}^{2P} e(\alpha x^4),$$
$$U(\alpha) = \sum_{h=1}^{U} e(\alpha u_h),$$

where $e(A)$ is an abbreviation for $e^{2\pi i A}$. Let

$$T^6(\alpha)U^2(\alpha) = \sum_m r_{14}(m)e(m\alpha).$$

In order to prove the main theorem, it suffices to prove that $r_{14}(N) > 0$. For any integer ξ, we have

(20) $$\int e(\alpha\xi)\,d\alpha = \begin{cases} 0 & \text{if } \xi \neq 0, \\ 1 & \text{if } \xi = 0, \end{cases}$$

where the range of integration is any interval of length 1. Hence

$$r_{14}(N) = \int T^6(\alpha)U^2(\alpha)e(-N\alpha)\,d\alpha. \tag{21}$$

Throughout the paper, a, q are subject to $a \leqq q$, $(a, q) = 1$. Let

$$S_{a,q} = \sum_{x=1}^{q} e_q(ax^4), \qquad \left(e_q(A) = e\left(\frac{A}{q}\right)\right).$$

For any real β, let

$$I(\beta) = \sum_{n=P^4}^{(2P)^4} \tfrac{1}{4}n^{-\frac{3}{4}}e(\beta n),$$

and let

$$T^*(\alpha, a, q) = q^{-1}S_{a,q}I\left(\alpha - \frac{a}{q}\right).$$

δ denotes a small positive number, fixed throughout the paper. Except in the next section, the constants implied by the symbol O depend only on δ, ϵ. C_1, C_2, C_3 denote positive absolute constants.

Approximations to $T(\alpha)$

The essence of the Hardy-Littlewood method consists in approximating to $T(\alpha)$ by $T^*(\alpha, a, q)$ when α lies in a certain interval surrounding the rational point a/q, and in using an upper bound for $T(\alpha)$ (such as that provided by Weyl's inequality) when α does not lie in any of these intervals. This apparatus of approximations and upper bounds is no simpler in the case of the exponent 4 than in the case of the general exponent. Consequently, we consider in this section the general sum

$$T(\alpha) = \sum_{x=P}^{2P} e(\alpha x^k),$$

where[5] $k \geqq 4$; and we allow the constants implied by the symbol O to depend on k as well as on δ, ϵ. The advantage of having the results set out in a form suitable for general reference may, perhaps, compensate for this intrusion of unnecessary generality.

We define

$$S_{a,q} = \sum_{x=1}^{q} e_q(ax^k),$$

$$I(\beta) = \sum_{n=P^k}^{(2P)^k} \frac{1}{k}n^{-1+1/k}e(\beta n),$$

$$T^*(\alpha, a, q) = q^{-1}S_{a,q}I\left(\alpha - \frac{a}{q}\right).$$

[5] The corresponding results for $k = 3$ (some of which are more precise) are given in the author's paper "On Waring's problem for cubes," *Acta Math.*

LEMMA 3. $S_{a,q} = O(q^{1-1/k})$.

PROOF. Landau, Satz 315.

LEMMA 4. *If* $|\beta| \leqq \frac{1}{2}$, *then*

$$I(\beta) = O(\min(P, P^{1-k}|\beta|^{-1})).$$

PROOF. The inequality $I(\beta) = O(P)$ is trivial. Also, if $|\beta| \leqq \frac{1}{2}$, we have

$$\sum_{n=n_1}^{n_2} e(\beta n) = O(|\beta|^{-1})$$

for any n_1, n_2. Hence, by Abel's Lemma,

$$\sum_{n=P^k}^{(2P)^k} n^{-1+1/k} e(\beta n) = O(P^{1-k}|\beta|^{-1}).$$

LEMMA 5. *If* $\alpha = \dfrac{a}{q} + \beta$, *where* $|\beta| \leqq \frac{1}{2}$, *then*

$$T^*(\alpha, a, q) = O(q^{-1/k} \min(P, P^{1-k}|\beta|^{-1})).$$

PROOF. Lemmas 3, 4.

LEMMA 6. *For any integer* ν, *let*

$$S_{a,q,\nu} = \sum_{x=1}^{q} e_q(ax^k + \nu x).$$

Then, if $\nu \neq 0$,

$$S_{a,q,\nu} = O(q^{\frac{1}{2}+\epsilon}(q, \nu)).$$

PROOF. Davenport-Heilbronn, Lemma 3.

LEMMA 7. *Suppose that*

$$H \geqq 1, \qquad q \leqq H^{1-\delta}, \qquad \beta = O(q^{-1}H^{1-k-\delta}),$$

and let $\nu \neq 0$ *be an integer. Then*

$$\int_0^H e\left(\beta\xi^k - \frac{\nu\xi}{q}\right) d\xi = -\frac{q}{2\pi i\nu}\left(e\left(\beta H^k - \frac{\nu H}{q}\right) - 1\right) + O(q\nu^{-2}H^{-\delta}).$$

PROOF. By integration by parts l times, the integral becomes

$$-\frac{q}{2\pi i\nu}\left(e\left(\beta H^k - \frac{\nu H}{q}\right) - 1\right) - \sum_{h=1}^{l-1}\left(\frac{q}{2\pi i\nu}\right)^{h+1}\left[e\left(-\frac{\nu\xi}{q}\right)D^h(e(\beta\xi^k))\right]_0^H$$
$$+\left(\frac{q}{2\pi i\nu}\right)^l \int_0^H e\left(-\frac{\nu\xi}{q}\right) D^l(e(\beta\xi^k))\, d\xi,$$

where D^h denotes the h^{th} derivative with respect to ξ, and $[F(\xi)]_0^H = F(H) - F(0)$. It is easily verified that

$$D^h(e(\beta\xi^k)) = \sum_{h/k \leqq r \leqq h} C(r, h, k)\beta^r \xi^{kr-h} e(\beta\xi^k),$$

where $C(r, h, k)$ depends only on the variables specified. For $0 \leqq \xi \leqq H$, $\frac{h}{k} \leqq r \leqq h$, we have

$$\begin{aligned}\beta^r \xi^{kr-h} &= O(q^{-r} H^{r(1-k-\delta)} H^{kr-h}) \\ &= O(q^{-h}(qH^{-1+\delta})^{h-r} H^{-h\delta}) \\ &= O(q^{-h} H^{-h\delta}).\end{aligned}$$

Hence, for $0 \leqq \xi \leqq H$,

$$D^h(e(\beta\xi^k)) = O(C(h, k) q^{-h} H^{-h\delta}).$$

Using this in the above expression, the second and third terms are

$$O\left(\sum_{h=1}^{l-1} \left(\frac{q}{|\nu|}\right)^{h+1} C(h, k) q^{-h} H^{-h\delta} + \left(\frac{q}{|\nu|}\right)^l HC(l, k) q^{-l} H^{-l\delta}\right).$$

Choose l to be the least integer for which $1 - l\delta \leqq -\delta$. Then the last expression is

$$O(q\nu^{-2} H^{-\delta}).$$

Lemma 8. *If $\alpha = \frac{a}{q} + \beta$, where $q \leqq P^{1-\delta}$ and $\beta = O(q^{-1}P^{1-k-\delta})$, then*

$$T(\alpha) = T^*(\alpha, a, q) + O(q^{\frac{1}{2}+\epsilon}).$$

Proof. We have

$$\begin{aligned}T(\alpha) &= \sum_{h=1}^{q} \sum_{(P-h)/q \leqq m \leqq (2P-h)/q} e\left(\left(\frac{a}{q} + \beta\right)(mq + h)^k\right) \\ &= \sum_{h=1}^{q} e_q(ah^k) \sum_{(P-h)/q \leqq m \leqq (2P-h)/q} e(\beta(mq + h)^k).\end{aligned}$$

By Poisson's summation formula,

$$\begin{aligned}\sum_{(P-h)/q \leqq m \leqq (2P-h)/q}' e(\beta(mq + h)^k) &= \int_{(P-h)/q}^{(2P-h)/q} e(\beta(\xi q + h)^k)\, d\xi \\ &\quad + \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} \int_{(P-h)/q}^{(2P-h)/q} e(\beta(\xi q + h)^k - \nu\xi)\, d\xi \\ &= q^{-1} \int_P^{2P} e(\beta\xi^k)\, d\xi + q^{-1} \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} e_q(\nu h) \int_P^{2P} e\left(\beta\xi^k - \frac{\nu\xi}{q}\right) d\xi,\end{aligned}$$

where $\sum'$ denotes that any term with $m = (P - h)/q$ or $m = (2P - h)/q$ is counted with a factor $\frac{1}{2}$. There are at most two values of h for which such a term exists, so the replacement of $\sum'$ by $\sum$ introduces only an error $O(1)$ in $T(\alpha)$. Also the first integral on the right is

$$\frac{1}{k} \int_{P^k}^{(2P)^k} \eta^{-1+1/k} e(\beta\eta)\, d\eta = \frac{1}{k} \sum_{n=P^k}^{(2P)^k} n^{-1+1/k} e(\beta n) + O(1),$$

since, for $\eta = n + O(1)$, $P^k \leqq n \leqq (2P)^k$, we have

$$\eta^{-1+1/k} e(\beta\eta) - n^{-1+1/k} e(\beta n) = O(n^{-1+1/k} \mid \beta \mid + n^{-2+1/k})$$
$$= O(P^{1-k} q^{-1} P^{1-k} + P^{1-2k})$$
$$= O(P^{-k}).$$

Hence

$$T(\alpha) = T^*(\alpha, a, q) + O(1) + \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} q^{-1} S_{a,q,\nu} \int_P^{2P} e\left(\beta\xi^k - \frac{\nu\xi}{q}\right) d\xi$$
$$= T^*(\alpha, a, q) + O(1) + \sum,$$

say.

The conditions of Lemma 7 are satisfied for $\int_0^P$ and $\int_0^{2P}$, hence

$$\sum = -\frac{1}{2\pi i} \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} \nu^{-1} S_{a,q,\nu} \left(e\left(\beta(2P)^k - \frac{2P\nu}{q}\right) - e\left(\beta P^k - \frac{P\nu}{q}\right)\right)$$
$$+ O\left(\sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} q^{-1} \mid S_{a,q,\nu} \mid q\nu^{-2} P^{-\delta}\right)$$
$$= -\frac{1}{2\pi i} \sum\nolimits_1 + \sum\nolimits_2,$$

say.

By Lemma 6,

$$\sum\nolimits_2 = O\left(P^{-\delta} \sum_{\nu=1}^{\infty} q^{\frac{1}{2}+\epsilon} \nu^{-2} (q, \nu)\right)$$
$$= O\left(q^{\frac{1}{2}+\epsilon} \sum_{d \mid q} d \sum_{m=1}^{\infty} (md)^{-2}\right)$$
$$= O(q^{\frac{1}{2}+2\epsilon}).$$

Also

$$\sum\nolimits_1 = e(\beta(2P)^k) \sum_{\mid \nu \mid > q^2} \nu^{-1} S_{a,q,\nu} e_q\left(-\frac{2P\nu}{q}\right)$$
$$- e(\beta P^k) \sum_{\mid \nu \mid > q^2} \nu^{-1} S_{a,q,\nu} e_q\left(-\frac{P\nu}{q}\right) + O\left(\sum_{\substack{\nu=-q^2 \\ \nu \neq 0}}^{q^2} \mid \nu \mid^{-1} \mid S_{a,q,\nu} \mid\right)$$
$$= e(\beta(2P)^k) \sum_{h=1}^{q} e_q(ah^k) \sum_{\nu=q^2+1}^{\infty} \frac{2i}{\nu} \sin \frac{\nu(h-2P)}{q} 2\pi$$
$$- e(\beta P^k) \sum_{h=1}^{q} e_q(ah^k) \sum_{\nu=q^2+1}^{\infty} \frac{2i}{\nu} \sin \frac{\nu(h-P)}{q} 2\pi + O\left(\sum_{\nu=1}^{q^2} \frac{1}{\nu} q^{\frac{1}{2}+\epsilon} (q, \nu)\right)$$

$$= O\left\{\sum_{h=1}^{q} \min\left(1, \frac{1}{q^2\left\|\frac{h-2P}{q}\right\|}\right) + \sum_{h=1}^{q} \min\left(1, \frac{1}{q^2\left\|\frac{h-P}{q}\right\|}\right) + q^{\frac{1}{2}+\epsilon} \sum_{d\mid q} d \sum_{m \leqq q^2/d} \frac{1}{md}\right\}$$

$$= O(q^{\frac{1}{2}+2\epsilon}),$$

where $\| \zeta \|$ denotes the distance of ζ from the nearest integer.

This proves Lemma 8.

Lemma 9. *If* $\alpha = \frac{a}{q} + \beta$, *where* $q \leqq P^{1-\delta}$ *and* $\beta = O(q^{-1}P^{1-k-\delta})$, *then*

$$T(\alpha) = O(q^{-1/k} \min (P, P^{1-k} \mid \beta \mid^{-1})).$$

Proof. The result follows from Lemmas 5, 8, since

$$q^{\frac{1}{2}+\epsilon} < q^{-1/k}P, \qquad q^{\frac{1}{2}+\epsilon} < q^{-1/k}P^{1-k} \mid \beta \mid^{-1}.$$

Lemma 10. (*Weyl's inequality*). *Let* $\kappa = \frac{1}{2^{k-1}}$. *Then*

$$\sum_{x=1}^{m} e_q(ax^k) = O(m^\epsilon q^\epsilon (m^{1-\kappa} + mq^{-\kappa} + m^{1-k\kappa}q^\kappa)).$$

Proof. Landau, Satz 267 ($K = 2^{k-1}$).

Lemma 11. *If* $P^{1-\delta} < q \leqq P^{k-1+\delta}$ *and* $\beta = O(q^{-1}P^{1-k-\delta})$, *then*

$$T(\alpha) = O(P^{1-\kappa+\delta}).$$

Proof. Let

$$S_m = \sum_{x^k \leqq m} e_q(ax^k).$$

By Lemma 10, if $m \leqq (2P)^k$,

$$S_m = O(P^\epsilon(P^{1-\kappa} + Pq^{-\kappa} + P^{1-k\kappa}q^\kappa))$$
$$= O(P^{1-\kappa+\kappa\delta+\epsilon}) = O(P^{1-\kappa+\delta}),$$

in virtue of the inequalities satisfied by q.

By partial summation,

$$T(\alpha) = \sum_{n=P^k}^{(2P)^k} (S_n - S_{n-1})e(\beta n)$$

$$= \sum_{n=P^k}^{(2P)^k} S_n(e(\beta n) - e(\beta(n+1))) + S_{(2P)^k}e(\beta((2P)^k + 1)) - S_{P^k-1}e(\beta P^k)$$

$$= O\left(P^{1-\kappa+\delta}\left(\sum_{n=P^k}^{(2P)^k} |\beta| + 1\right)\right)$$

$$= O(P^{1-\kappa+\delta} + P^{1-\kappa+\delta}P^k q^{-1} P^{1-k-\delta})$$
$$= O(P^{1-\kappa+\delta}).$$

The Farey dissection

Let $Q = [P^{3+\delta}]$. For $q \leqq P^{\frac{1}{2}}$, let $\mathfrak{M}_{a,q}$ denote the interval

$$\left|\alpha - \frac{a}{q}\right| \leqq \frac{1}{qQ}. \tag{22}$$

Two intervals $\mathfrak{M}_{a_1,q_1}$, $\mathfrak{M}_{a_2,q_2}$ corresponding to different pairs a, q do not overlap, since

$$\left|\frac{a_1}{q_1} - \frac{a_2}{q_2}\right| \geqq \frac{1}{q_1 q_2} > \frac{1}{q_1 q_2}\frac{2P^{\frac{1}{2}}}{[P^{3+\delta}]} \geqq \frac{1}{Q}\left(\frac{1}{q_1} + \frac{1}{q_2}\right).$$

By Lemmas 5, 8, 9, (with $k = 4$), if α is in $\mathfrak{M}_{a,q}$ and $\alpha = \dfrac{a}{q} + \beta$,

$$T(\alpha),\ T^*(\alpha, a, q) = O(q^{-\frac{1}{4}} \min (P, P^{-3} \mid \beta \mid^{-1})), \tag{23}$$

$$T(\alpha) - T^*(\alpha, a, q) = O(q^{\frac{1}{2}+\epsilon}). \tag{24}$$

It is well known[6] that for any real α there exist A, q (where A is an integer) such that

$$\left|\alpha - \frac{A}{q}\right| \leqq \frac{1}{qQ}, \qquad q \leqq Q - 1, \qquad (A, q) = 1. \tag{25}$$

Suppose that α satisfies

$$\frac{1}{Q} < \alpha < 1 + \frac{1}{Q}. \tag{26}$$

Then

$$A \geqq q\alpha - \frac{1}{Q} > \frac{q}{Q} - \frac{1}{Q} \geqq 0,$$

$$A \leqq q\alpha + \frac{1}{Q} < q + \frac{q}{Q} + \frac{1}{Q} \leqq q + 1.$$

Hence $1 \leqq A \leqq q$, and so we are entitled to use the letter a instead of A in (25).

If $q \leqq P^{\frac{1}{2}}$, the points α satisfying (25) form precisely $\mathfrak{M}_{a,q}$. We denote by $\mathfrak{m}$ all points of the interval (26) which do not belong to any $\mathfrak{M}_{a,q}$. It follows that for any α in $\mathfrak{m}$, there exist a, q such that

$$\left|\alpha - \frac{a}{q}\right| \leqq \frac{1}{qQ}, \qquad P^{\frac{1}{2}} < q \leqq Q - 1. \tag{27}$$

---

[6] Landau, Satz 158.

By (21), taking the interval of integration to be (26), we have

$$(28)\quad r_{14}(N) = \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_{\mathfrak{M}_{a,q}} T^6(\alpha)U^2(\alpha)e(-N\alpha)\,d\alpha + \int_{\mathfrak{m}} T^6(\alpha)U^2(\alpha)e(-N\alpha)\,d\alpha.$$

Major arcs

LEMMA 12.

$$\sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_{\mathfrak{M}_{a,q}} |\, T^6(\alpha) - T^{*6}(\alpha, a, q)\,|\,|\,U(\alpha)\,|^2\,d\alpha = O(U^2P^{2-\frac{1}{4}+\epsilon}).$$

PROOF. By (23), (24), if α is in $\mathfrak{M}_{a,q}$ and $\alpha = \dfrac{a}{q} + \beta$,

$$T^6(\alpha) - T^{*6}(\alpha, a, q) = O(q^{\frac{3}{4}+\epsilon}q^{-5/4}\min(P^5, P^{-15}\,|\,\beta\,|^{-5})).$$

Hence

$$\int_{\mathfrak{M}_{a,q}} |\, T^6(\alpha) - T^{*6}(\alpha, a, q)|\,d\alpha = O\left(q^{-\frac{1}{2}+\epsilon}\int_0^{\infty}\min(P^5, P^{-15}\beta^{-5})\,d\beta\right)$$
$$= O(q^{-\frac{1}{2}+\epsilon}P).$$

Thus the sum of the lemma is

$$O(U^2\sum_{q \leqq P^{\frac{1}{2}}}\sum_{a} q^{-\frac{1}{2}+\epsilon}P) = O(U^2P\sum_{q \leqq P^{\frac{1}{2}}} q^{\frac{1}{2}+\epsilon})$$
$$= O(U^2P^{1+\frac{1}{2}(\frac{3}{2}+\epsilon)}).$$

LEMMA 13. *If $\overline{\mathfrak{M}}_{a,q}$ denotes the part of the interval* (26) *not belonging to $\mathfrak{M}_{a,q}$, then*

$$\sum_{q \leqq P^{\frac{1}{2}}}\sum_{a}\int_{\overline{\mathfrak{M}}_{a,q}} |\, T^*(\alpha, a, q)|^6\,|\,U(\alpha)|^2\,d\alpha = O(U^2).$$

PROOF. Since the integrand is a periodic function of α with period 1, we can take the range of integration to be $q^{-1}Q^{-1} \leqq |\,\beta\,| \leqq \frac{1}{2}$, where $\alpha = \dfrac{a}{q} + \beta$. By Lemma 5, with $k = 4$, the sum of the lemma is

$$O\left(U^2\sum_{q \leqq P^{\frac{1}{2}}}\sum_{a}\int_{q^{-1}Q^{-1}}^{\infty} q^{-6/4}P^{-18}\beta^{-6}\,d\beta\right) = O(U^2\sum_{q \leqq P^{\frac{1}{2}}}\sum_{a} q^{-3/2}P^{-18}q^5P^{5(3+\delta)})$$
$$= O(U^2P^{-3+5\delta}\sum_{q \leqq P^{\frac{1}{2}}} q^{9/2})$$
$$= O(U^2P^{-3+5\delta+11/4}),$$

whence the result.

Minor arcs

LEMMA 14. *In $\mathfrak{m}$, $T(\alpha) = O(P^{\frac{1}{2}+\delta})$.*

PROOF. We have seen that for any α in $\mathfrak{m}$ there exist a, q satisfying (27).

CASE 1. $P^{1/2} < q \leqq P^{1-\delta}$. By Lemma 9, with $k = 4$,

$$T(\alpha) = O(q^{-1/4}P) = O(P^{7/8}).$$

CASE 2. $P^{1-\delta} < q \leqq Q - 1 < P^{3+\delta}$. Lemma 11 with $k = 4$.

LEMMA 15. $\int_{\mathfrak{m}} | T(\alpha) |^6 | U(\alpha) |^2 \, d\alpha = O(U^2 P^{2-\frac{1}{2}+3\mu+5\delta})$.

PROOF. By Lemma 14 we have, in $\mathfrak{m}$,

$$T^4(\alpha) = O(P^{4-\frac{1}{2}+4\delta}). \tag{29}$$

By (20),

$$\int | T(\alpha)U(\alpha) |^2 \, d\alpha,$$

taken over any interval of length 1, is precisely the number of solutions of (2) subject to (3). By (17), (18), (19), the conditions of Lemma 1 are satisfied. Hence

$$\int | T(\alpha)U(\alpha) |^2 \, d\alpha = O(P^2 U^2 P^{3\mu-4+2\epsilon}). \tag{30}$$

The assertion of the lemma follows from (29), (30).

The singular series

LEMMA 16. $\int_0^1 T^{*6}(\alpha, a, q)e(-n\alpha) \, d\alpha = q^{-6}(S_{a,q})^6 e_q(-na)R(n)$, *where, for* $N - P^4 < n < N$, $R(n)$ *satisfies*

$$2^{-30}P^2 < R(n) < 2^8 P^2. \tag{31}$$

PROOF. By (20), the integral in question is

$$q^{-6}(S_{a,q})^6 e_q(-na) \sum_{n_1,\cdots,n_6} \frac{1}{4^6} (n_1 \cdots n_6)^{-\frac{3}{4}},$$

where the variables of summation are subject to

$$P^4 \leqq n_1, \cdots, n_6 \leqq (2P)^4, \qquad n_1 + \cdots + n_6 = n. \tag{32}$$

Since $n_1, \cdots, n_5$ determine n_6 uniquely, we have

$$R(n) < \frac{1}{4^6} (2P)^{20} P^{-18} = 2^8 P^2.$$

Also, by (16), $100P^4 \leqq N < 101P^4$, so that $99P^4 < n < 101P^4$. Hence, if $18P^4 \leqq n_1, \cdots, n_5 \leqq 19P^4$, the value of n_6 determined by the last condition in (32) satisfies also the other conditions.

It follows that

$$R(n) > \frac{1}{4^6} P^{20} (2P)^{-18} = 2^{-30} P^2.$$

This establishes Lemma 16.

Let

$$A(n, q) = \sum_a q^{-6}(S_{a,q})^6 e_q(-na).$$

LEMMA 17. *$A(n, q) = O(q^{-\frac{1}{2}})$.*

PROOF. Lemma 3.

LEMMA 18. *If $(q_1, q_2) = 1$ then $A(n, q_1q_2) = A(n, q_1)A(n, q_2)$.*

PROOF. Landau, Satz 282.

The following notation corresponds to that of Landau, pp. 280–302, when $k = 4$, $s = 6$, except that in some cases we are precluded from using the same symbols.

For any prime p, we define $\gamma = 1$ if $p > 2$ and $\gamma = 4$ if $p = 2$.

For any prime p, and any l, n, let $N(p^l, n)$ denote the number of solutions of

$$x_1^4 + \cdots + x_6^4 \equiv n \pmod{p^l}, \qquad 1 \leqq x_i \leqq p^l,$$

in which not all of $x_1, \cdots, x_6$ are divisible by p.

LEMMA 19. *Let $4\rho + \sigma$ be the exact power to which p divides n, where $0 \leqq \sigma \leqq 3$. Let*

$$l_0 = \max(4\rho + \sigma + 1, 4\rho + \gamma).$$

Then

$$A(n, p^l) = 0 \text{ for } l > l_0,$$

and

$$\chi_p(n) = \sum_{\nu=0}^{\infty} A(n, p^\nu) = p^{-5\gamma}N(p^\gamma, 0) \sum_{\tau=0}^{\rho-1} p^{-2\tau} + p^{-2\rho-5\gamma} N\left(p^\gamma, \frac{n}{p^{4\rho}}\right),$$

where, if $\rho = 0$, the sum on the right is to be read as zero.

PROOF. This is the case $k = 4$, $s = 6$ of Landau's Satz 293. (Note that Landau uses P for p^γ and β instead of ρ.)

COROLLARY. *If $p \nmid 2n$, then $A(n, p^l) = 0$ for $l > 1$.*

LEMMA 20. *If $p > 2$, then $N(p^\gamma, n) > 0$ for all n.*

PROOF. By Landau, Sätze 300 and 301, with $s = 6$, it suffices to prove that

$$6 \geqq \frac{p^\gamma - 1}{p - 1}(4, p - 1) + 1.$$

For $p > 2$, the right-hand side is $(4, p - 1) + 1 \leqq 5$.

LEMMA 21. *If $p = 2$, and $n \equiv 1, 2, 3, 4, 5,$ or $6 \pmod{16}$, then $N(p^\gamma, n) > 0$.*

PROOF. Since $p^\gamma = 16$, the result is obvious.

LEMMA 22. *For any prime p, and any $n \equiv 1, 2, 3, 4, 5,$ or $6 \pmod{16}$,*

$$\chi_p(n) \geqq p^{-20}.$$

PROOF. CASE 1. Suppose $p^4 \nmid n$, so that $\rho = 0$. By Lemmas 19, 20, 21,

$$\begin{aligned}\chi_p(n) &= p^{-5\gamma}N(p^\gamma, n)\\ &\geqq p^{-5\gamma}\\ &\geqq p^{-20}.\end{aligned}$$

CASE 2. Suppose $p^4|n$, so that $\rho \geqq 1$ and $p > 2$. By Lemmas 19, 20,

$$\chi_p(n) \geqq p^{-5\gamma}N(p^\gamma, 0) \geqq p^{-5}.$$

LEMMA 23. *For any prime p and any n, $|A(n, p)| < C_1p^{-2}$.*

PROOF. Landau, Satz 318, with $s = 6$.

LEMMA 24. *For any prime p and any n, $\chi_p(n) > 1 - C_2p^{-2}$.*

PROOF. Landau, Satz 322, with $s = 6$.

LEMMA 25. *If $n \equiv 1, 2, 3, 4, 5,$ or 6 (mod 16), the series*

$$\mathfrak{S}(n) = \sum_{q=1}^{\infty} A(n, q)$$

converges absolutely, and

$$\mathfrak{S}(n) > C_3.$$

PROOF. By the Corollary to Lemma 19, and by Lemma 23, if $p \nmid 2n$,

$$|\chi_p(n) - 1| = |A(n, p)| < C_1p^{-2}.$$

Hence the product $\prod_p \chi_p(n)$ is absolutely convergent, therefore by Lemma 18 so is the series $\sum_{q=1}^{\infty} A(n, q)$, and the two are equal.

By Lemmas 22, 24,

$$\begin{aligned}\prod_p \chi_p(n) &> \sum_{p \leqq C_2} p^{-20} \prod_{p > C_2} (1 - C_2p^{-2})\\ &= C_3.\end{aligned}$$

LEMMA 26. *For $\eta \geqq 1$, $\sum_{q \geqq \eta} |A(n, q)| = O(n^\epsilon \eta^{-\frac{1}{4}})$.*

PROOF. Any positive integer q is representable as $2^\tau q_1q_2$, where

(1) q_1 is *quadratfrei* and odd;

(2) q_2 is odd and composed of prime powers with exponents $\geqq 2$;

(3) $(q_1, q_2) = 1$.

By Lemma 17,

$$A(n, 2^\tau) = O(2^{-\frac{1}{4}\tau}), \qquad A(n, q_2) = O(q_2^{-\frac{1}{4}}).$$

Also, by Lemma 19, if $p^l | q_2$, $p^{l+1} \nmid q_2$, and $A(n, p^l) \neq 0$, then $l \leqq l_0 = 4\rho + \sigma + 1$, where $4\rho + \sigma$ is the exact power to which p divides n. Since $l \geqq 2$, it follows that $p^l | p^{2(l-1)} | n^2$. Hence, if $A(n, q_2) \neq 0$, we must have $q_2 | n^2$.

By Lemmas 18, 23,

$$|A(n, q_1)| = \prod_{p \mid q_1} |A(n, p)| < \prod_{p \mid q_1} (C_1 p^{-2}) = O(q_1^{-2+\epsilon}).$$

Hence

$$\sum_{q \geqq \eta} |A(n, q)| = O\left(\sum_{q_1} \sum_{\substack{q_2 \\ q_2 \mid n^2 \\ 2^r q_1 q_2 \geqq \eta}} \sum_{r} q_1^{-2+\epsilon} q_2^{-\frac{1}{2}} 2^{-\frac{1}{2}r}\right)$$

$$= O\left(\eta^{-\frac{1}{4}} \sum_{q_1=1}^{\infty} \sum_{q_2 \mid n^2} \sum_{r=0}^{\infty} q_1^{-7/4+\epsilon} q_2^{-\frac{1}{4}} 2^{-\frac{1}{4}r}\right)$$

$$= O(\eta^{-\frac{1}{4}} n^{\epsilon}).$$

Proof of the Theorem

By (28),

$$r_{14}(N) = \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_0^1 T^{*6}(\alpha, a, q) U^2(\alpha) e(-N\alpha)\, d\alpha$$

$$- \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_{\overline{\mathfrak{M}}_{a,q}} T^{*6}(\alpha, a, q) U^2(\alpha) e(-N\alpha)\, d\alpha$$

$$+ \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_{\mathfrak{M}_{a,q}} (T^6(\alpha) - T^{*6}(\alpha, a, q)) U^2(\alpha) e(-N\alpha)\, d\alpha$$

$$+ \int_{\mathfrak{m}} T^6(\alpha) U^2(\alpha) e(-N\alpha)\, d\alpha.$$

By Lemma 13, the second sum is $O(U^2)$. By Lemma 12, the third sum is $O(U^2 P^{2-\frac{1}{4}+\epsilon})$. By Lemma 15, the final integral is $O(U^2 P^{2-\frac{1}{4}+3\mu+5\delta})$. By (17), the last of these three error terms is the largest. Also the first sum, by Lemma 16, is

$$\sum_{h=1}^{U} \sum_{j=1}^{U} \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} \int_0^1 T^{*6}(\alpha, a, q) e(u_h \alpha + u_j \alpha - N\alpha)\, d\alpha$$

$$= \sum_{h=1}^{U} \sum_{j=1}^{U} \sum_{q \leqq P^{\frac{1}{2}}} \sum_{a} q^{-6} (S_{a,q})^6 e_q(-(N - u_h - u_j)a) R(N - u_h - u_j)$$

$$= \sum_{h=1}^{U} \sum_{j=1}^{U} \sum_{q \leqq P^{\frac{1}{2}}} A(N - u_h - u_j, q) R(N - u_h - u_j).$$

By Lemma 26,

$$\sum_{q \leqq P^{\frac{1}{2}}} A(N - u_h - u_j, q) = \mathfrak{S}(N - u_h - u_j) + O(P^{-\frac{1}{8}+\epsilon}).$$

Also, since $u_h, u_j < P^{\mu+3} = o(P^4)$, $n = N - u_h - u_j$ satisfies the condition of Lemma 16. Thus in the above sum, $R(N - u_h - u_j)$ satisfies

(33) $$2^{-30} P^2 < R(N - u_h - u_j) < 2^8 P^2.$$

Hence

$$r_{14}(N) = \sum_{h=1}^{U} \sum_{j=1}^{U} \mathfrak{S}(N - u_h - u_j)R(N - u_h - u_j) + O(U^2 P^{-\frac{1}{4}+\epsilon} P^2)$$
$$+ O(U^2 P^{2-\frac{1}{4}+3\mu+5\delta}).$$

By (15), $N - u_h - u_j \equiv 1, 2, 3, 4, 5,$ or $6 \pmod{16}$. Hence, by Lemma 25, $\mathfrak{S}(N - u_h - u_j) > C_3$. Thus the above sum is greater than

$$U^2 C_3 2^{-30} P^2.$$

By (17),

$$3\mu - \frac{1}{2} = \frac{729}{1567} \quad \frac{1}{2} < 0.$$

Hence

$$r_{14}(N) > 0.$$

VICTORIA UNIVERSITY,
MANCHESTER, ENGLAND.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

L'ASPECT QUALITATIF DE LA THÉORIE ANALYTIQUE DES POLYNOMES

par J. Dieudonné

(Received March 21, 1939)

1. La plupart des problèmes qui se posent dans ce que j'ai proposé récemment[1] d'appeler la *théorie analytique des polynomes*, rentrent dans le schéma général suivant: on considère une famille $\mathcal{F}$ de polynomes d'une variable complexe, de degré *borné*, et qui dépendent d'un certain nombre de *paramètres* variables; il s'agit de savoir *s'il existe*, et dans l'affirmative de *déterminer*, des régions du plan complexe (non identiques au plan tout entier) telles que *tout* polynome de la famille possède, dans une telle région, *un nombre de zéros au moins égal à un nombre donné* $r > 0$.

La formulation même d'un tel problème montre qu'il se décompose naturellement en deux parties: la première, qu'on peut appeler la partie *qualitative* du problème, consiste à déterminer les nombres r pour lesquels on peut affirmer l'*existence* de régions ayant la propriété désirée; cela fait, on doit passer à la partie *quantitative* du problème, c'est-à-dire *déterminer* ces régions pour chacune des valeurs de r où leur existence est assurée.

Je me propose, dans cet article, de montrer comment, dans tous les cas envisagés jusqu'ici, la partie *qualitative* du problème précédent peut, par l'emploi d'une méthode uniforme, se ramener à un autre problème dont la difficulté est sensiblement moindre dans la plupart des cas. Cette méthode s'appuie sur des propriétés de *compacité* (ou de *familles normales*, pour employer la terminologie de M. P. Montel); elle n'est d'ailleurs pas essentiellement nouvelle, et on la retrouve çà et là dans la plupart des travaux sur la théorie analytique des polynomes.[2] Mais on ne semble pas avoir remarqué avec quelle facilité elle s'applique à de nombreux problèmes résolus jusqu'à présent par des moyens très particuliers et très divers, et souvent moyennant des hypothèses superflues.

Ce fait accentue encore le contraste entre l'aspect qualitatif et l'aspect quantitatif de la Théorie analytique des polynomes, car les résultats quantitatifs obtenus jusqu'ici l'ont tous été au moyen d'artifices variés ne se rattachant à aucune idée générale.

---

[1] *J. Dieudonné, La théorie analytique des polynomes d'une variable* (Mémorial des Sciences mathématiques, fasc. 93). Dans ce qui suit, nous nous référons à cet opuscule en le désignant simplement par "*Mémorial.*"

[2] Voir en particulier *P. Montel, Sur quelques limites pour les modules des zéros des polynomes* (Comment. Math. Helv., t. 7, 1934–35, p. 178–200), et *J. Dieudonné, Sur la variation des zéros des dérivées des fractions rationnelles* (Ann. de l'Ecole Normale Supérieure, 3e série, t. 54, 1937, p. 101–150); dans ce dernier travail, la méthode générale exposée ici est appliquée à l'étude qualitative complète d'un problème particulier.

2. Précisons d'abord le problème énoncé ci-dessus. Soit N le degré *maximum* des polynomes de la famille $\mathcal{F}$; nous ferons la convention suivante: tout polynome de la famille $\mathcal{F}$ dont le degré n sera inférieur à N sera considéré comme ayant un zéro multiple d'ordre $N - n$ au point à l'infini. Nous supposons en outre que la famille $\mathcal{F}$ ne contient pas le polynome identiquement nul: tout polynome de $\mathcal{F}$ possède alors *exactement N zéros* (chacun compté avec son ordre de multiplicité) dans le plan *fermé* par l'adjonction du point à l'infini.

Comme il s'agit d'étudier la répartition des zéros, on peut supposer que, si la famille $\mathcal{F}$ contient un polynome $P(x)$, elle contient aussi le polynome $aP(x)$, quelle que soit la constante complexe $a \neq 0$.

Si $P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_Nx^N$, on peut faire correspondre à ces polynomes le point de *l'espace projectif complexe à N dimensions* $\boldsymbol{P}^N$ dont les coordonnées homogènes sont $(a_0, a_1, a_2, \cdots, a_N)$. A la famille $\mathcal{F}$ correspond ainsi un sous-ensemble F de $\boldsymbol{P}^N$, de sorte que tous les polynomes de $\mathcal{F}$ correspondant à un même point de F ont les mêmes zéros.

Dans l'énoncé du problème général, nous avons parlé de *régions* du plan contenant toujours r zéros au moins de tout polynome de la famille $\mathcal{F}$; précisons maintenant que nous entendons par là des *sous-ensembles fermés* du plan complexe *fermé* par adjonction du point à l'infini. Dans ces conditions, on peut aussi, en vertu de la continuité des racines d'une équation en fonction des paramètres, supposer *fermée* la famille $\mathcal{F}$, autrement dit supposer que F est un ensemble *fermé* dans $\boldsymbol{P}^N$; d'où, en vertu de la *compacité* de l'espace projectif complexe, résulte que F est *compact*.

3. Ces préliminaires étant posés, nous cherchons donc les valeurs de l'entier r telles qu'il existe un ensemble fermé E_r, non identique au plan tout entier, et contenant r zéros au moins de *tout* polynome de la famille $\mathcal{F}$; il est clair que, si r possède cette propriété, tout entier $r' < r$ la possède également; il suffit donc de chercher le *maximum* ρ des nombres r pour lesquels la propriété précédente est vraie. Nous désignerons cette recherche sous le nom de *problème global*.

Pour le résoudre, étudions d'abord le problème qui se pose dans les mêmes termes, avec la seule différence qu'on impose en outre aux ensembles E_r la condition de *ne pas contenir un point donné* x_0 du plan. Désignons par $\sigma(x_0)$ le *maximum* des nombres r pour lesquels il existe des ensembles E_r satisfaisant à cette condition supplémentaire; nous appellerons *problème local* (relatif au point x_0) la recherche de ce nombre. Il est clair qu'on a ensuite $\rho = \operatorname{Max} \sigma(x_0)$ lorsque x_0 parcourt le plan fermé; la résolution du problème local pour tout point du plan entraîne donc celle du problème global.

Or, le nombre $\sigma(x_0)$ peut encore être défini par les deux conditions suivantes:

1° quel que soit le voisinage V de x_0, il existe un polynome de la famille $\mathcal{F}$ possédant *au moins* $N - \sigma(x_0)$ zéros dans V;

2° il existe un voisinage V_0 de x_0 tel que *tout* polynome de la famille $\mathcal{F}$ possède *au plus* $N - \sigma(x_0)$ zéros dans V_0.

Désignons alors par $\lambda(x_0)$ *l'ordre de multiplicité maximum* du point x_0 comme

zéro des polynomes de la famille $\mathcal{F}$ ($\lambda(x_0) = 0$ si aucun polynome de $\mathcal{F}$ ne s'annule en x_0). Nous allons voir qu'on a

$$\sigma(x_0) = N - \lambda(x_0). \tag{1}$$

En effet, comme il existe au moins un polynome de $\mathcal{F}$ ayant au point x_0 un zéro multiple d'ordre $\lambda(x_0)$, donc ayant au moins $\lambda(x_0)$ zéros dans le voisinage V_0, on a

$$\lambda(x_0) \leqq N - \sigma(x_0) \tag{2}$$

D'autre part, soit (V_i) $(i = 1, 2, \cdots)$ une suite de voisinages de x_0 dont l'intersection se réduise au point x_0 ; à chaque indice i correspond un polynome $P_i(x)$ de la famille $\mathcal{F}$ possédant au moins $N - \sigma(x_0)$ zéros dans V_i. Comme F est *compact*, on peut extraire de la suite $(P_i(x))$ une suite partielle convergeant vers un polynome $P_0(x)$ de $\mathcal{F}$, et, d'après la continuité des racines, $P_0(x)$ possède un zéro multiple d'ordre $N - \sigma(x_0)$ au moins au point x_0, d'où

$$N - \sigma(x_0) \leqq \lambda(x_0) \tag{3}$$

La comparaison de (2) et (3) donne la relation (1). On en déduit que, si μ est le *minimum* des nombres $\lambda(x_0)$ lorsque x_0 parcourt le plan fermé, on a

$$\rho = N - \mu. \tag{4}$$

4. Pour appliquer les résultats précédents, il est essentiel que F soit *fermé*; s'il n'en est pas ainsi, il faut commencer par remplacer F par son *adhérence* $\bar{F}$, c'est-à-dire adjoindre à F les points d'accumulation de toutes les suites de points de F, qui n'appartiennent pas à F. Dans le calcul de $\lambda(x_0)$, il faudra faire entrer en ligne de compte, non seulement les polynomes de la famille $\mathcal{F}$ donnée, mais aussi ceux qui correspondent aux points de $\bar{F} - F$.

Par exemple, la famille $\mathcal{F}$ est donnée le plus souvent par l'expression des coefficients $a_0, a_1, \cdots, a_N$ en fonction de *paramètres* indépendants dont chacun est assujetti à décrire un certain sous-ensemble du plan complexe; si les sous-ensembles où varient certains de ces paramètres *ne sont pas fermés* (dans le plan fermé par adjonction du point à l'infini), il faudra adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre chacun des paramètres considérés vers un point d'accumulation quelconque de l'ensemble où il varie. C'est ainsi que, souvent, certains paramètres peuvent prendre toutes les valeurs complexes *finies*; il faut alors adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre ces paramètres vers le point à l'infini. Par exemple, à la famille des polynomes de la forme $1 + x + ax^2 + bx^3$ où a et b prennent toutes les valeurs complexes finies, il faut adjoindre tous les polynomes de la forme $ax^2 + bx^3$, où a et b sont encore des nombres finis arbitraires.

Considérons de même la famille $\mathcal{F}$ formée des numérateurs des dérivées $m^{\text{èmes}}$ non identiquement nulles de fractions rationnelles de degré n (m et n

donnés) dont chacun des pôles et des zéros décrit un sous-ensemble fermé donné; il peut exister certaines positions de ces points pour lesquelles la fraction se réduit à un polynome de degré inférieur à m, et ces positions doivent par suite être exclues, puisque la dérivée m^e est alors identiquement nulle; mais il faut adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre les zéros et les pôles des fractions rationnelles considérées vers les positions singulières précédentes; et les polynomes correspondant à ces points d'accumulation n'appartiennent pas nécessairement à la famille $\mathcal{F}$ initiale.[3]

5. Donnons maintenant quelques exemples d'application de la méthode que nous venons de décrire.

Considérons en premier lieu la famille des polynomes

$$1 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{5}$$

de degré n donné, dont certains des coefficients $a_1, \cdots, a_n$ sont *fixes*, les autres pouvant prendre toutes les valeurs complexes *finies*; la détermination du nombre $\sigma(\infty)$, et des ensembles bornés E_r contenant au moins r zéros de chacun de ces polynomes (pour chaque $r \leqq \sigma(\infty)$) constitue ce qu'on appelle le *problème de Landau-Montel.* D'après ce qu'on a vu plus haut, il faut tout d'abord adjoindre aux polynomes (5) ceux qu'on obtient en remplaçant par 0 les coefficients de (5) qui sont fixés; on obtient ainsi une famille fermée. Soit alors a_p le *premier* coefficient de (5) qui soit variable, a_q le *dernier* coefficient *donné et différent de* 0; $\sigma(\infty)$ est le degré *minimum* des polynomes de la famille considérée (*fermée* comme il vient d'être dit); or, il est immédiat que ce degré est *le plus petit des nombres* p, q, ce qui résout qualitativement le problème de Landau-Montel.[4]

De façon générale, pour une famille *fermée* $\mathcal{F}$ de polynomes, le nombre $\sigma(\infty)$ sera le *degré minimum* des polynomes de $\mathcal{F}$.

6. Considérons maintenant la famille des polynomes de la forme

$$a_1P_1(x) + a_2P_2(x) + \cdots + a_kP_k(x) \tag{6}$$

où $a_1, \cdots, a_k$ peuvent prendre toutes les valeurs complexes finies, et où $P_i(x)$ est un polynome de degré n ($n \geqq k$), dont tous les zéros sont assujettis à varier dans un *domaine circulaire fermé* C_i ($i = 1, 2, \cdots, k$). En outre, nous supposerons d'abord que les domaines circulaires $C_1, C_2, \cdots, C_k$ n'ont *aucun point commun deux à deux.*

Cette famille est fermée; nous allons montrer qu'on a $\rho = n - k + 1$. En effet, il existe des points n'appartenant à aucun des C_i; soit x_0 l'un d'eux; il nous suffira de montrer que $\lambda(x_0) = k - 1$ (car en un point y_0 appartenant à un

[3] On montre cependant que ce sont encore des numérateurs de dérivées $m^{\text{ièmes}}$ de fractions rationnelles de degré n (voir le mémoire de l'auteur cité dans la note 2, p. 110, lemme 1).

[4] Ce raisonnement ne diffère que par la forme de celui de M. Montel dans le mémoire cité plus haut.

zéro des polynomes de la famille $\mathcal{F}$ ($\lambda(x_0) = 0$ si aucun polynome de $\mathcal{F}$ ne s'annule en x_0). Nous allons voir qu'on a

$$\sigma(x_0) = N - \lambda(x_0). \tag{1}$$

En effet, comme il existe au moins un polynome de $\mathcal{F}$ ayant au point x_0 un zéro multiple d'ordre $\lambda(x_0)$, donc ayant au moins $\lambda(x_0)$ zéros dans le voisinage V_0, on a

$$\lambda(x_0) \leqq N - \sigma(x_0) \tag{2}$$

D'autre part, soit (V_i) $(i = 1, 2, \cdots)$ une suite de voisinages de x_0 dont l'intersection se réduise au point x_0 ; à chaque indice i correspond un polynome $P_i(x)$ de la famille $\mathcal{F}$ possédant au moins $N - \sigma(x_0)$ zéros dans V_i. Comme F est *compact*, on peut extraire de la suite $(P_i(x))$ une suite partielle convergeant vers un polynome $P_0(x)$ de $\mathcal{F}$, et, d'après la continuité des racines, $P_0(x)$ possède un zéro multiple d'ordre $N - \sigma(x_0)$ au moins au point x_0, d'où

$$N - \sigma(x_0) \leqq \lambda(x_0) \tag{3}$$

La comparaison de (2) et (3) donne la relation (1). On en déduit que, si μ est le *minimum* des nombres $\lambda(x_0)$ lorsque x_0 parcourt le plan fermé, on a

$$\rho = N - \mu. \tag{4}$$

4. Pour appliquer les résultats précédents, il est essentiel que F soit *fermé*; s'il n'en est pas ainsi, il faut commencer par remplacer F par son *adhérence* $\bar{F}$, c'est-à-dire adjoindre à F les points d'accumulation de toutes les suites de points de F, qui n'appartiennent pas à F. Dans le calcul de $\lambda(x_0)$, il faudra faire entrer en ligne de compte, non seulement les polynomes de la famille $\mathcal{F}$ donnée, mais aussi ceux qui correspondent aux points de $\bar{F} - F$.

Par exemple, la famille $\mathcal{F}$ est donnée le plus souvent par l'expression des coefficients $a_0, a_1, \cdots, a_N$ en fonction de *paramètres* indépendants dont chacun est assujetti à décrire un certain sous-ensemble du plan complexe; si les sous-ensembles où varient certains de ces paramètres *ne sont pas fermés* (dans le plan fermé par adjonction du point à l'infini), il faudra adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre chacun des paramètres considérés vers un point d'accumulation quelconque de l'ensemble où il varie. C'est ainsi que, souvent, certains paramètres peuvent prendre toutes les valeurs complexes *finies*; il faut alors adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre ces paramètres vers le point à l'infini. Par exemple, à la famille des polynomes de la forme $1 + x + ax^2 + bx^3$ où a et b prennent toutes les valeurs complexes finies, il faut adjoindre tous les polynomes de la forme $ax^2 + bx^3$, où a et b sont encore des nombres finis arbitraires.

Considérons de même la famille $\mathcal{F}$ formée des numérateurs des dérivées $m^{\text{èmes}}$ non identiquement nulles de fractions rationnelles de degré n (m et n

donnés) dont chacun des pôles et des zéros décrit un sous-ensemble fermé donné; il peut exister certaines positions de ces points pour lesquelles la fraction se réduit à un polynome de degré inférieur à m, et ces positions doivent par suite être exclues, puisque la dérivée m^e est alors identiquement nulle; mais il faut adjoindre à l'ensemble F les points d'accumulation des suites qu'on obtient en faisant tendre les zéros et les pôles des fractions rationnelles considérées vers les positions singulières précédentes; et les polynomes correspondant à ces points d'accumulation n'appartiennent pas nécessairement à la famille $\mathcal{F}$ initiale.[3]

5. Donnons maintenant quelques exemples d'application de la méthode que nous venons de décrire.

Considérons en premier lieu la famille des polynomes

$$1 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{5}$$

de degré n donné, dont certains des coefficients $a_1, \cdots, a_n$ sont *fixes*, les autres pouvant prendre toutes les valeurs complexes *finies*; la détermination du nombre $\sigma(\infty)$, et des ensembles bornés E_r contenant au moins r zéros de chacun de ces polynomes (pour chaque $r \leqq \sigma(\infty)$) constitue ce qu'on appelle le *problème de Landau-Montel.* D'après ce qu'on a vu plus haut, il faut tout d'abord adjoindre aux polynomes (5) ceux qu'on obtient en remplaçant par 0 les coefficients de (5) qui sont fixés; on obtient ainsi une famille fermée. Soit alors a_p le *premier* coefficient de (5) qui soit variable, a_q le *dernier* coefficient *donné et différent de* 0; $\sigma(\infty)$ est le degré *minimum* des polynomes de la famille considérée (*fermée* comme il vient d'être dit); or, il est immédiat que ce degré est *le plus petit des nombres* p, q, ce qui résout qualitativement le problème de Landau-Montel.[4]

De façon générale, pour une famille *fermée* $\mathcal{F}$ de polynomes, le nombre $\sigma(\infty)$ sera le *degré minimum* des polynomes de $\mathcal{F}$.

6. Considérons maintenant la famille des polynomes de la forme

$$a_1P_1(x) + a_2P_2(x) + \cdots + a_kP_k(x) \tag{6}$$

où $a_1, \cdots, a_k$ peuvent prendre toutes les valeurs complexes finies, et où $P_i(x)$ est un polynome de degré n ($n \geqq k$), dont tous les zéros sont assujettis à varier dans un *domaine circulaire fermé* C_i ($i = 1, 2, \cdots, k$). En outre, nous supposerons d'abord que les domaines circulaires $C_1, C_2, \cdots, C_k$ n'ont *aucun point commun deux à deux*.

Cette famille est fermée; nous allons montrer qu'on a $\rho = n - k + 1$. En effet, il existe des points n'appartenant à aucun des C_i; soit x_0 l'un d'eux; il nous suffira de montrer que $\lambda(x_0) = k - 1$ (car en un point y_0 appartenant à un

[3] On montre cependant que ce sont encore des numérateurs de dérivées $m^{\text{ièmes}}$ de fractions rationnelles de degré n (voir le mémoire de l'auteur cité dans la note 2, p. 110, lemme 1).

[4] Ce raisonnement ne diffère que par la forme de celui de M. Montel dans le mémoire cité plus haut.

des C_i, on a évidemment $\lambda(y_0) = n$). Cela revient à voir que le déterminant

$$\begin{vmatrix} P_1(x_0) & P_2(x_0) & \cdots & P_k(x_0) \\ P_1'(x_0) & P_2'(x_0) & \cdots & P_k'(x_0) \\ \cdots & \cdots & \cdots & \cdots \\ P_1^{(k-1)}(x_0) & \cdots & \cdots & P_k^{(k-1)}(x_0) \end{vmatrix}$$

n'est pas nul. Or ce déterminant est *linéaire* par rapport à chacune des racines des $P_i(x)$, considérées comme variables; d'après le théorème de Grace,[5] si ce déterminant était nul, il existerait, dans chacun des domaines C_i, un point z_i tel que

$$\begin{vmatrix} (x_0 - z_1)^n & (x_0 - z_2)^n & \cdots & (x_0 - z_k)^n \\ (x_0 - z_1)^{n-1} & (x_0 - z_2)^{n-1} & \cdots & (x_0 - z_k)^{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ (x_0 - z_1)^{n-k+1} & (x_0 - z_2)^{n-k+1} & \cdots & (x_0 - z_k)^{n-k+1} \end{vmatrix} = 0$$

Or ce déterminant est égal au produit des facteurs $(x_0 - z_i)^{n-k+1}$ $(i = 1, 2, \cdots, k)$ par le déterminant de Vandermonde des quantités $(x_0 - z_i)$, et il résulte des hypothèses faites qu'aucun de ces facteurs ne peut être nul, d'où la proposition.

L'hypothèse faite sur les C_i est essentielle; montrons en effet que si deux domaines circulaires C_i, C_j ont *un point intérieur commun* z_0, on a $\rho = 0$. Prenons en effet $P_i(x) = (x - z_0)^n$, et soit a un nombre assez petit pour que le polynome $(x - z_0)^n + a(x - x_0)^n$ ait toutes ses racines dans C_j, ce qui est possible d'après l'hypothèse; en prenant $P_j(x)$ égal à ce polynome, $a_i = 1$, $a_j = -1$, et $a_h = 0$ pour les indices h différents de i et j, le polynome (6) a un zéro multiple d'ordre n au point x_0, d'où la proposition.

7. Soit $\mathcal{F}_n$ la famille des polynomes de la forme

$$(x - x_1)^{m+1}(x - x_2)^{m+1} \cdots (x - x_n)^{m+1}\left[\frac{a_1}{(x - x_1)^{m+1}} + \cdots + \frac{a_q}{(x - x_q)^{m+1}} + \cdots + \frac{a_n}{(x - x_n)^{m+1}}\right] \tag{7}$$

où $(a_1, a_2, \cdots, a_n)$ décrit, dans l'espace projectif complexe P^{n-1} un ensemble fermé G de sorte qu'il existe un nombre $k > 0$ tel que, pour tout choix d'indices $i_1, i_2, \cdots, i_p$ parmi les nombres $1, 2, \cdots, n$, on ait

$$|a_{i_1} + a_{i_2} + \cdots + a_{i_p}| \geqq k \cdot \operatorname*{Max}_{1 \leqq r \leqq p} |a_{i_r}| \tag{8}$$

C'est le cas par exemple lorsque tous les a_i ne prennent que des valeurs *positives ou nulles*.

En outre, $x_{q+1}, x_{q+2}, \cdots, x_n$ peuvent prendre toutes les valeurs complexes

[5] Voir *Mémorial*, p. 12, théorème VIIc.

finies; enfin, $x_1, x_2, \cdots, x_q$ varient arbitrairement dans un ensemble *borné* fermé E.

On peut évidemment supposer que $\operatorname{Max}_{1 \leqq i \leqq n} |a_i| = 1$; d'après (8) le coefficient de la plus haute puissance de x dans le polynome (7) a son module compris entre k et n; lorsque les x_i varient en restant bornés, et que le point $(a_1, a_2, \cdots, a_n)$ varie arbitrairement dans G, on n'obtient, comme limites de suites de polynomes de $\mathfrak{F}_n$, que des polynomes de cette famille.

Pour *fermer* $\mathfrak{F}_n$, il suffit donc de considérer le cas où un certain nombre de points x_i $(q + 1 \leqq i \leqq n)$ tendent vers le point à l'infini, les autres restant bornés et le point $(a_1, a_2, \cdots, a_n)$ variant arbitrairement dans G; si, parmi les indices i pour lesquels x_i reste borné, il en existe un tel que a_i ne tende pas vers 0, on voit immédiatement qu'on obtient, comme limites, des polynomes des familles $\mathfrak{F}_r$ définies comme $\mathfrak{F}_n$, mais pour un nombre r de termes compris entre q et n. Si on n'est pas dans ce cas, les limites sont, soit des polynomes des familles $\mathfrak{F}_r$, soit des polynomes qu'on obtient en ajoutant une *constante arbitraire* dans le crochet des formules analogues à (7) définissant les polynomes des familles $\mathfrak{F}_r$ pour $r < n$.

Cela étant, la condition (8) montre immédiatement que, dans les polynomes de la famille fermée ainsi obtenue, le terme de plus haut degré est au moins de degré $(m + 1)(q - 1)$, cette borne inférieure étant effectivement atteinte lorsqu'on laisse fixes les a_i et qu'on fait tendre $x_{q+1}, \cdots, x_n$ vers le point à l'infini. On a donc

$$\sigma(\infty) = (m + 1)(q - 1) \tag{9}$$

ce qui généralise un théorème de M. Fekete.[6]

8. Un raisonnement analogue montrerait que si on considère encore les polynomes de la forme (7), avec les mêmes hypothèses sur les x_i, mais où les a_i peuvent prendre toutes les valeurs complexes finies, on a cette fois[7]

$$\sigma(\infty) = m(q - 1) \tag{10}$$

d'où, par transformation homographique, on déduit aisément que l'on a aussi $\rho = m(q - 1)$ pour cette famille de polynomes.

En particulier, si $x_1, x_2, \cdots, x_n$ sont *fixes* et *distincts*, les a_i variant arbitrairement, on a $\rho = m(n - 1)$. Par un raisonnement de continuité facile, on étend ce résultat aux polynomes de la forme

$$P_1(x)P_2(x) \cdots P_n(x) \left[\frac{a_1}{P_1(x)} + \frac{a_2}{P_2(x)} + \cdots + \frac{a_n}{P_n(x)} \right] \tag{11}$$

où $P_i(x)$ est un polynome de degré $m + 1$, dont tous les zéros varient dans un voisinage *suffisamment petit* d'un point x_i, les points x_i étant tous *distincts*

[6] *Mémorial*, p. 60, théorème XLVIII.

[7] *J. Dieudonné*, loc.cit. (note 2), p. 146.

$(i = 1, 2, \cdots, n)$ et les a_i arbitraires. Mais on n'a pas ici de résultat aussi simple que pour les polynomes de la forme (6), en ce qui concerne la détermination des domaines où peuvent varier les zéros des $P_i(x)$ pour que le résultat précédent demeure valable. On peut seulement montrer que $\rho = 0$ si on n'impose *aucune condition* à ces domaines: il suffit de prendre $P_i(x) = x^{m+1} - b_i$, le b_i étant distincts, et de considérer la décomposition en éléments simples de la fraction

$$\frac{1}{(y - b_1)(y - b_2) \cdots (y - b_n)}$$

en y remplaçant y par x^{m+1}, pour voir que le polynome (11) peut se réduire à une constante non nulle.

Signalons encore deux résultats analogues à celui du n°7. Si, dans les polynomes (11), les zéros de chacun des $P_i(x)$ varient dans un ensemble fermé *borné*, et si de plus les a_i satisfont à la condition

$$|a_1 + a_2 + \cdots + a_n| \geqq k \cdot \operatorname*{Max}_{1 \leqq i \leqq n} |a_i|$$

on a $\sigma(\infty) = (m + 1)(n - 1)$ pour ces polynomes, ce qui généralise qualitativement un résultat de M. M. J. v. sz. Nagy et Marden.[8]

Si maintenant on considère les polynomes de la forme (11), où $P_i(x) = (x - x_i)^{p_i}$, et où on fait sur les a_i et les x_i les *mêmes* hypothèses qu'au n°7, le même raisonnement que dans ce n° montre que

$$\sigma(\infty) = p_1 + p_2 + \cdots + p_q - \operatorname{Max}(p_1, p_2, \cdots, p_q)$$

formule qui redonne bien (9) quand les p_i sont tous égaux à $m + 1$.

NANCY, FRANCE

[8] *Mémorial*, p. 55, théorème XLII.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

STRUCTURE AND AUTOMORPHISMS OF SEMI-SIMPLE LIE GROUPS IN THE LARGE[1]

BY N. JACOBSON

(Received May 3, 1939)

The present paper attempts to fill in several gaps in the literature relating Lie algebras to Lie groups. It is well known that there is a (1–1) correspondence between Lie subalgebras of the Lie algebra $\mathfrak{L}$ of $\mathfrak{G}$ and its closed local subgroups. However in general different closed local subgroups may generate the same closed subgroup of $\mathfrak{G}$. We can show nevertheless that $\mathfrak{G}$ is semi-simple (simple) if and only if $\mathfrak{L}$ is semi-simple (simple). We also give a new set of linear groups which represent the classes of locally isomorphic simple Lie groups and which is somewhat simpler than Cartan's original list.[2] Omitting a finite number of exceptions these are merely the important geometric linear groups with real complex and quaternionic elements (unimodular, orthogonal and symplectic[3] groups). We determine the group $\mathfrak{A}$ of bicontinuous automorphisms of these "unexceptional" groups and discuss the structure of $\mathfrak{A}/\mathfrak{I}$, $\mathfrak{I}$ the set of inner automorphisms.

1. We begin with a brief resumé of the local theory. A topological space U is a *local group*[4] (group germ, group nucleus) if it contains an open set V in which a composition xy in U is defined such that

1. If x, y, z, xy, yz are in V then $x(yz) = (xy)z$.
2. There is a point 1 in V such that $1x = x1 = x$ for any x in V.
3. For each x in V there is an x^{-1} in V such that $xx^{-1} = x^{-1}x = 1$.
4. xy and x^{-1} are continuous in x and y.

U is a *local Lie group* if it is an r-cell and in place of 4. we have

4'. xy and x^{-1} are analytic in x and y in the sense that the coördinates are analytic functions of those of x and y.

W is a local subgroup of U if it is a local group relative to the same composition and topology as defined in U. Two local subgroups are regarded as identical if their intersection is open in each of the local groups. W is invariant if for each neighborhood W_1 of 1 in W there exist symmetric[5] neighborhoods N and Z of the identity in U and W respectively such that $x^{-1}zx \in W_1$ if $x \in N$ and $z \in Z$. Two

[1] Presented to the National Academy of Sciences Oct. 24, 1938.

[2] Cartan [2].

[3] This term has been introduced by Prof. Weyl (The classical groups, Princeton 1939) in place of the earlier and undesirable terms *Abelian* or *complex* linear groups.

[4] Sometimes called *group germ* or *group nucleus*. The present term has been adopted in the forthcoming translation of Pontrjagin's book on continuous groups.

[5] W_1 is symmetric if W_1^{-1} the set of points w_1^{-1}, w_1 in W_1, $= W_1$.

local groups U and U' are isomorphic if there exists a homeomorphism $x \to x'$ between suitable symmetric neighborhoods N and N' of 1 and $1'$ such that if x, y, xy are in N then $x'y' = (xy)'$ is in N'. It follows that $1 \to 1'$, $x^{-1} \to (x')^{-1}$. We may also suppose that if $x', y', x'y'$ are in N' then xy is in N.[6]

Suppose now that U is a local Lie group and let M be a cell neighborhood of 1 in which canonical parameters[7] are defined. Then if $x = (\xi_1, \cdots, \xi_r) \in N$ a concentric sphere of half the radius of M then x^2 has coördinates $2\xi_i$ and x has a unique square root $x^{\frac{1}{2}}$ in N. We remark that x may not have a unique n^{th} root $(n > 2)$ in N but it has a unique *proper* n^{th} root y in the sense that $y, y^2, \cdots y^{n-1}$ are all in N. For our purpose the existence and uniqueness of square roots suffices. We define $x^{\frac{1}{4}} = (x^{\frac{1}{2}})^{\frac{1}{2}}, \cdots, x^{p/2^m} = x^{(p-1)/2^m} x^{1/2^m}$ if $0 \leqq p \leqq 2^m$ and $x^{-p/2^m} = (x^{p/2^m})^{-1}$. We define x^α for $-1 \leqq \alpha \leqq 1$ by a limiting process. Finally if $|\alpha| > 1$ but $(\alpha\xi_1, \cdots, \alpha\xi_r)$ is still in N we define x^α as the element y in N such that $y^{1/\alpha} = x$. The canonical coördinates of x^α are $\alpha\xi_i$. Since the usual rules for powers hold the elements x^α, x fixed form a 1-dimensional local subgroup and these local subgroups cover the neighborhood N without overlapping.

It has been shown by G. Birkhoff[8] that the Lie algebra associated with U may be introduced in the following way: If x and y are in a suitable neighborhood N' of 1 then $\lim_{\alpha\to 0} (x^\alpha y^\alpha)^{1/\alpha} = x + y$ and $\lim_{\alpha \to 0} (x^\alpha y^\alpha x^{-\alpha} y^{-\alpha})^{1/\alpha^2} = [x, y]$ exist. The coördinates of $x + y$ are $\xi_i + \eta_i$ if $x = (\xi_1, \cdots, \xi_r)$, $y = (\eta_1, \cdots, \eta_r)$. $[x, y]$ is bilinear in x and y and satisfies

$$[x, y] = -[y, x] \quad [x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$$

($-x$ denotes the elements with coördinates $-\xi_i$, $x + (-x) = 0 \equiv 1$). We may regard N' as the interior of a sphere in r-dimensional Cartesian space $\mathfrak{L}$ and $x + y$ as vector addition in this space. If x and y are arbitrary in $\mathfrak{L}$ we may choose $\rho \neq 0$ so that $\rho x, \rho y \in N'$ and define $[x, y] = \rho^{-2}[\rho x, \rho y]$. This is independent of ρ and together with the operation $+$ defines $\mathfrak{L}$ as a Lie algebra over the field of real numbers. We call $\mathfrak{L}$ *the Lie algebra* of U. The operations in $\mathfrak{L}$ have been defined by algebraic and topological processes in U. Conversely we may express the group operation by

$$xy = x + y + \tfrac{1}{2}[x, y] + \cdots$$

where all the terms are pure commutators. It follows that any isomorphism between local Lie groups induces an isomorphism between their Lie algebras and conversely.

It has been shown by E. Cartan that if W is a closed local subgroup of U there is a neighborhood of 1 in W whose elements all belong to a subalgebra $\mathfrak{M}$ of $\mathfrak{L}$.[9]

---

[6] If we choose N_1 so that $N_1^2 \subset N$ and the corresponding N_1' then this will hold for N_1 in place of N.

[7] Cf. Eisenhart [1], p. 44 or Birkhoff [1], p. 73.

[8] Birkhoff [1]. Cf. also Pontrjagin's book. For the definitions needed from the theory of Lie algebras see Jacobson [1], p. 875.

[9] Cartan [3], p. 22.

$\mathfrak{M}$ may be defined as the set of all the limiting directions of elements of W converging to 1. Thus W is a local Lie group with $\mathfrak{M}$ as its Lie algebra. $\mathfrak{M} = 0$ if and only if W is discrete. It follows also that any totally disconnected closed local subgroup of U is discrete. W is invariant if and only if $\mathfrak{M}$ is an ideal.

2. If U is a neighborhood of 1 in a topological group $\mathfrak{G}$ it is a local group relative to the multiplication and topology defined in $\mathfrak{G}$. We recall that $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are locally isomorphic if they have isomorphic local groups U_1 and U_2. $\mathfrak{G}$ is a *Lie group* if U can be taken as a local Lie group. It follows that the Lie groups $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are locally isomorphic if and only if their Lie algebras (the Lie algebras of the local Lie groups U_1 and U_2) $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are isomorphic. If $\mathfrak{G}_1$ is connected and simply connected Schreier has shown that if U_1 is any neighborhood of 1 in $\mathfrak{G}_1$ there exists a neighborhood $V_1 \subset U_1$ such that the elements of $\mathfrak{G}_1$ are representable as $v_1v_2 \cdots v_m$, v_i in V_1 and $v_1v_2 \cdots v_m = 1$ if and only if this relation can be reduced to $1 = 1$ by a sequence of identifications $uv = w$ where u, v, w are in U_1.[10] It follows that if $\mathfrak{G}_2$ is connected any local isomorphism between $\mathfrak{G}_1$ and $\mathfrak{G}_2$ is obtained from a unique continuous homomorphism of $\mathfrak{G}_1$ into $\mathfrak{G}_2$. The group $\mathfrak{D}$ mapped into 1 in this mapping is discrete. If $\mathfrak{G}_2$ is simply connected also, $\mathfrak{D} = 1$ and the mapping is an isomorphism. In particular if $\mathfrak{G}_1 = \mathfrak{G}_2 = \mathfrak{G}$ any local automorphism corresponds to a (bicontinuous) automorphism of $\mathfrak{G}$. If $\mathfrak{G}$ is not simply connected there may exist local automorphism which do not generate automorphisms in the large. On the other hand any automorphism determines a local automorphism of $\mathfrak{G}$ and if $\mathfrak{G}$ is connected distinct automorphisms will define distinct local automorphisms. Since the latter induce linear transformations in $\mathfrak{L}$ it is natural to topologize the group $\mathfrak{A}$ of automorphism using the usual Euclidean topology in the space of linear transformations.

Any closed subgroup $\mathfrak{H}$ of $\mathfrak{G}$ determines a closed local subgroup W of U and hence a subalgebra $\mathfrak{M}$ of $\mathfrak{L}$. Hence $\mathfrak{H}$ is a Lie group. If $\mathfrak{H}$ is invariant so is W and hence $\mathfrak{M}$ is an ideal. If $\mathfrak{M}$ is any subalgebra of $\mathfrak{L}$ the closed local group W determined by $\mathfrak{M}$ may not be a neighborhood of 1 of any closed subgroup of $\mathfrak{G}$.[11] If $\mathfrak{H}^*$ is the smallest closed subgroup containing W, W^* its local group and $\mathfrak{M}^*$ its Lie algebra then $W^* \geqq W$ and $\mathfrak{M}^* \geqq \mathfrak{M}$. If $\mathfrak{M}$ is an ideal W is an invariant local group and $\mathfrak{H}^*$ is invariant in $\mathfrak{G}_1$ the component of 1 in $\mathfrak{G}$. Hence $\mathfrak{M}^*$ is an ideal. Similarly if $\mathfrak{M}$ is commutative so is $\mathfrak{M}^*$. Since the elements $(g^\alpha h^\alpha g^{-\alpha} h^{-\alpha})^{1/\alpha^2}$ converge in the direction through $[g, h]$ it follows from Cartan's theorem that the Lie algebra of the derived group[12] $\mathfrak{G}'$ contains the derived algebra $\mathfrak{L}'$.

We shall call a topological group $\mathfrak{G}$ topologically semi-simple or briefly *t-semi-simple* if every closed commutative invariant subgroup of $\mathfrak{G}$ is discrete. If $\mathfrak{G}_1$ is

[10] Schreier [1], p. 25.

[11] A piece of a geodesic in an irrational direction on the torus is an example of this type.

[12] The smallest closed subgroup containing all the commutators $g_1^{-1}g_2^{-1}g_1g_2$.

locally isomorphic to $\mathfrak{G}$, $\mathfrak{H}_1$ a closed commutative invariant subgroup of $\mathfrak{G}_1$ then $\mathfrak{H}_1$ is discrete. For, otherwise $\mathfrak{H}_1$ determines a local subgroup of this type and hence by the local isomorphism a commutative invariant local subgroup not discrete in $\mathfrak{G}$. If $\mathfrak{G}$ is connected and $\mathfrak{D}$ a discrete invariant subgroup, $\mathfrak{D} \leqq \mathfrak{C}$ the center of $\mathfrak{G}$.[13] Hence if $\mathfrak{D}$ is a commutative invariant subgroup its closure $\overline{\mathfrak{D}}$ is discrete and $\overline{\mathfrak{D}} = \mathfrak{D} \leqq \mathfrak{C}$ also. The natural homomorphism between $\mathfrak{G}$ and $\mathfrak{G}/\mathfrak{C}$ is a local isomorphism. Hence $\mathfrak{G}/\mathfrak{C}$ is t-semi-simple. The group $\mathfrak{C}_1 \geqq \mathfrak{C}$ which corresponds to the center of $\mathfrak{G}/\mathfrak{C}$ is discrete and invariant.[14] Thus $\mathfrak{C}_1 = \mathfrak{C}$, $\mathfrak{G}/\mathfrak{C}$ has no commutative invariant subgroup $\neq 1$ and hence no discrete invariant subgroup $\neq 1$.

$\mathfrak{G}$ is *t-simple* if every closed invariant subgroup of $\mathfrak{G}$ is either open or discrete. It follows that $\mathfrak{G}$ is t-semi-simple unless its component of 1 is commutative. If $\mathfrak{G}$ is connected its closed invariant subgroups $\neq \mathfrak{G}$ are discrete. Then either $\mathfrak{G} = \mathfrak{C}$ or the latter is discrete. Furthermore $\mathfrak{G}/\mathfrak{C}$ is t-simple and has no discrete invariant subgroup $\neq 1$. Hence $\mathfrak{G}/\mathfrak{C}$ is simple in the usual sense. We suppose henceforth that if $\mathfrak{G}$ is t-simple it is also t-semi-simple.

Now suppose that $\mathfrak{G}$ is a t-semi-simple Lie group, $\mathfrak{L}$ its Lie algebra. $\mathfrak{L}$ is semi-simple. For if $\mathfrak{M}$ is a commutative ideal, $\mathfrak{H}^*$ the smallest closed subgroup containing $\mathfrak{M}$ is commutative and invariant. The converse that $\mathfrak{G}$ is t-semi-simple if $\mathfrak{L}$ is semi-simple is trivial. If $\mathfrak{G}$ is t-simple (and t-semi-simple) $\mathfrak{L}$ is semi-simple and if not simple, $\mathfrak{L} = \mathfrak{L}_1 \oplus \mathfrak{L}_2$[15] where $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are proper ideals. $\mathfrak{L}_i$ generates $\mathfrak{G}_1$ the component of 1 but since the elements of $\mathfrak{L}_1$ and $\mathfrak{L}_2$ commute, $\mathfrak{G}_1$ is commutative contrary to assumption. Conversely if $\mathfrak{L}$ is simple, $\mathfrak{G}$ is t-simple.

Suppose again that $\mathfrak{G}$ is t-semi-simple and $\mathfrak{L}_1$ is an ideal in $\mathfrak{L}$, $\mathfrak{G}_1$ the closed invariant subgroup determined by $\mathfrak{L}_1$ and $\mathfrak{L}_1^*$ the Lie algebra of $\mathfrak{G}_1$. Since $\mathfrak{L}_1^*$ is an ideal in $\mathfrak{L}$ it is semi-simple and hence $\mathfrak{L}_1^* = \mathfrak{L}_1 \oplus \mathfrak{M}$. The elements of $\mathfrak{M}$ commute with those of $\mathfrak{L}_1$ and hence they generate a subgroup of the center of $\mathfrak{G}_1$. It follows that $\mathfrak{M} = 0$, $\mathfrak{L}_1^* = \mathfrak{L}_1$. Hence in this case we have a (1–1) correspondence between closed connected invariant subgroups of $\mathfrak{G}$ and ideals of $\mathfrak{L}$. Suppose $\mathfrak{G}$ is connected and $\mathfrak{L} = \mathfrak{L}_1 \oplus \cdots \oplus \mathfrak{L}_s$, $\mathfrak{L}_i$ simple and $\mathfrak{G}_i$ the group generated by $\mathfrak{L}_i$. Then $\mathfrak{G} = \mathfrak{G}_1\mathfrak{G}_2 \cdots \mathfrak{G}_s$. The component of 1 of $\mathfrak{D}_i = \mathfrak{G}_i \cap (\mathfrak{G}_1 \cdots \mathfrak{G}_{i-1}\mathfrak{G}_{i+1} \cdots \mathfrak{G}_s)$ corresponds to $\mathfrak{L}_i \cap (\mathfrak{L}_1 + \cdots + \mathfrak{L}_{i+1} + \mathfrak{L}_{i+1} + \cdots + \mathfrak{L}_s) = 0$ and hence consists of the point 1. Thus $\mathfrak{D}_i$ is discrete and is contained in $\mathfrak{C}$ the center. If $\mathfrak{C} = 1$, $\mathfrak{G}$ is a direct product of simple groups. In the general case $\mathfrak{G}/\mathfrak{C}$ is a direct product of simple groups. It has been shown by Cartan[16] that the inner automorphisms $(\mathfrak{G}/\mathfrak{C})$ form the component of 1 in

---

[13] If $d \in \mathfrak{D}$ then the set of elements $g^{-1}dg$, $g \in \mathfrak{G}$ is connected since $g \to g^{-1}dg$ is a continuous mapping. Hence $g^{-1}dg$ consists of a single point. i.e. $g^{-1}dg = 1$.

[14] If $\mathfrak{H}/\mathfrak{C}$ and $\mathfrak{C}$ are discrete then $\mathfrak{H}$ is discrete. For $\mathfrak{C}$ the inverse image of the open set 1 is open in $\mathfrak{H}$. Since $\mathfrak{C}$ is discrete any point contained in $\mathfrak{C}$ is also open in $\mathfrak{H}$ and hence $\mathfrak{H}$ is discrete.

[15] Cartan [1], p. 52.

[16] Cartan [4], p. 8.

the group of automorphisms $\mathfrak{A}$. Thus this component has no commutative invariant subgroup and if $\mathfrak{G}$ is t-simple the component is simple.

3. Let $\mathfrak{R}$ be a complete valued associative ring with an identity[17] i.e. there is defined a real valued function $|x|$ for x in $\mathfrak{R}$ such that

1. $|x| \geqq 0, = 0$ if and only if $x = 0$,
2. $|x| = |-x|$,
3. $|x+y| \leqq |x| + |y|$,
4. $|xy| \leqq |x||y|$,
5. $\mathfrak{A}$ is complete in the topology defined by the metric $d(x, y) = |x - y|$.

The set of units u in $\mathfrak{R}$ form a group $\mathfrak{U}$. Since the product is continuous in $\mathfrak{R}$ by 3 and 4 it is continuous in the subspace $\mathfrak{U}$. If $|x| < 1$ it follows as usual that $y = 1 + x + x^2 + \cdots$ exists and $(1 - x)y = y(1 - x) = 1$. Also for a given $\epsilon > 0$ there exists $\delta > 0$ such that $|x| < \delta$ implies $|y - 1| < \epsilon$. Hence $\mathfrak{U}$ is open in $\mathfrak{R}$ and x^{-1} is continuous at $x = 1$. Now suppose N is any open set containing a^{-1} then aN is open and contains 1 since $x \to xa$ is a homeomorphism. There exists a neighborhood M of 1 such that $M^{-1} \leqq aN$. Hence $N' = Ma$ is an open set about a such that $(N')^{-1} \leqq N$. Thus x^{-1} is continuous at every point and $\mathfrak{U}$ is a topological group.

We suppose now that $\mathfrak{R}$ is a simple algebra over R the field of real numbers, i.e. $\mathfrak{R} = R_n, C_n$ or Q_n the set of $n \times n$ matrices with elements in the real, complex or quaternionic fields. If $x = (\xi_{ij}) = \sum \xi_{ij} e_{ij}$, e_{ij} a matric basis set $|x| = (\sum \xi_{ij}\bar{\xi}_{ij})^{\frac{1}{2}}$. Then $\mathfrak{R}$ is a complete valued ring.[18] According as $\mathfrak{R} = R_n, C_n$ or Q_n we write $\mathfrak{U} = L(R, n), L(C, n)$ or $L(Q, n)$. Using the methods introduced by v. Neumann[19] we can define $\exp x = 1 + x + \dfrac{x^2}{2!} + \cdots$ for any x and $\log x = (x - 1) - \dfrac{(x-1)^2}{2} +$ for $|x - 1| < 1$. These functions are continuous and

$$\exp(\log x) = x \quad \text{if} \quad |x - 1| < 1$$

$$\log(\exp x) = x \quad \text{if} \quad |x| < \log 2.$$

Thus the mapping $x \to \exp x$ is a homeomorphism between the neighborhood $|x| < \log 2$ of 0 and a neighborhood M of 1. Since $\exp x \exp(-x) = 1$ the latter is in $\mathfrak{U}$. It is readily seen that $\mathfrak{U}$ is a Lie group. Since

$$(\exp x)(\exp y) = \exp(x + y)$$

if $xy = yx$ we have $(\exp x)^2 = \exp 2x$ and hence the real coördinates of ξ_{ij} in $\exp x$, $x = (\xi_{ij})$ are canonical. Hence if N is the image of $|x| < \delta_1$, $\delta_1 \leqq \frac{1}{2}\log 2$ under the mapping $x \to a = \exp x$ then a has the unique square root $a^{\frac{1}{2}} = \exp \frac{1}{2}x$.

[17] Deuring [1], p. 93.
[18] von Neumann [1], p. 6. The proof given there holds also for R_n and Q_n.
[19] von Neumann [1].

As in §2 we have a^α defined topologically-algebraically and $a^\alpha = \exp(\alpha x)$. Since

$$|\exp x - \exp y| = O|x - y| \qquad |\log x - \log y| = O|x - y|$$

in our neighborhoods we can prove for $b = \exp y$

$$(a^\alpha b^\alpha)^{1/\alpha} = \exp\left[\frac{1}{\alpha}\log(\exp \alpha x \exp \alpha y)\right] \rightarrow \exp(x + y)$$

and

$$(a^\alpha b^\alpha a^{-\alpha} b^{-\alpha})^{1/\alpha^2} \rightarrow \exp[x, y] \qquad [x, y] = xy - yx$$

if $|x|, |y| < \delta_2$.[20] Thus the correspondence $x \rightarrow \exp x = a$ sets up an isomorphism between the Lie algebra of $\mathfrak{U}$ and the Lie algebra $\mathfrak{R}$ in which $x + y$ is as in the associative algebra and $[x, y] = xy - yx$.

If $\mathfrak{G}$ is a closed subgroup of $\mathfrak{U}$ the result of Cartan's quoted above shows that $\mathfrak{G}$ has a neighborhood of 1 consisting of the matrices $\exp x$ where $|x| < \rho$ and $x \in \mathfrak{L}$ a Lie subalgebra of $\mathfrak{R}$.

4. If $\mathfrak{U} = L(R, n)$ or $L(C, n)$ it is well known[21] that the derived groups $L'(R, n) = L_1(R, n)$ the set of matrices of determinant 1 and $L'(C, n) = L_1(C, n)$. Since $\det(\exp x) = \exp(\operatorname{tr} x)$, $\operatorname{tr} x = \sum \xi_{ii}$ if $x = (\xi_{ij})$ the Lie algebras of $L'(R, n)$ and $L'(C, n)$ are respectively the derived algebras R_n' and C_n'. Q_n may be represented by $2n \times 2n$ complex matrices. We define tr a, det a as the trace and determinant in this representation. As is well-known tr a, det $a \in R$. It is readily seen that Q_n' consists of the elements of trace 0. We show below (§8) that det $a > 0$ for a in $L(Q, n)$ and $L'(Q, n)$ consists of the elements of determinant 1. Hence we have, as before, that the Lie algebra of $L'(Q, n)$ is Q_n'. R_n', $C_n'(n > 1)$, Q_n' are simple and hence $L'(R, n)$, $L'(C, n)$, $L'(Q, n)$ are t-simple Lie groups.[22] We call these the real, complex and quaternionic unimodular groups.

If S is an automorphism in $\mathfrak{R}$[23] it induces an automorphism in $\mathfrak{U}'$ and in the Lie algebra $\mathfrak{R}'$. If $a = \exp x$ by the continuity of S we have $a^S = \exp x^S$ and hence S in $\mathfrak{R}'$ induces S in $\mathfrak{U}'$. If S is an anti-automorphism in $\mathfrak{R}$, $a \rightarrow (a^S)^{-1}$ and $x \rightarrow -x^S$ are automorphisms in $\mathfrak{U}'$ and $\mathfrak{R}'$ respectively. Since $(a^S)^{-1} = \exp(-x^S)$, $x \rightarrow -x^S$ generates the automorphism in $\mathfrak{U}'$. It has been shown that every automorphism in $\mathfrak{R}'$ is of one of these two types. Applying this we obtain the following groups $\mathfrak{A}$ of automorphisms for the unimodular groups.

1. $L'(R, n)$. The automorphism of R_n are $x \rightarrow s^{-1}xs$, the anti-automorphisms

---

[20] Cf. Birkhoff [1], p. 78.

[21] A proof of this theorem is given in §8.

[22] The results on Lie algebras required here and in the rest of this section may be found in Jacobson [2], pp. 545-548.

[23] We mean here an automorphism in the algebra $\mathfrak{R}$, i.e. $(x\alpha)^S = x^S\alpha$ for real α. This is equivalent to the condition that S be a continuous automorphism in the ring $\mathfrak{R}$.

are $x \to s^{-1}x's$ where x' is the transposed of x. These give the automorphisms $a \to s^{-1}as$ and $a \to s^{-1}(a')^{-1}s$ in $L'(R, n)$. The former set form a subgroup $\cong L(R, n)/R^*$, R^* the set of matrices $\alpha 1 \neq 0$ and this subgroup has index 1 or 2 according as $n = 2$ or $n > 2$. If n is even $R^* \leqq L^+(R, n)$ the set of matrices of positive determinant and $L^+(R, n)/R^*$ has index 2 in $L(R, n)/R^*$. Since any matrix in $L^+(R, n)$ has the form αa_1, $\alpha \in R^*$ and $a_1 \in L'(R, n)$ we may choose two elements a_1, $-a_1$ in each class mod R^*. Hence $L^+(R, n)/R^* \cong L'(R, n)/D \cong \mathfrak{J}$, $D = (1, -1)$, $\mathfrak{J}$ the group of inner automorphisms. If n is odd every class mod R^* contains matrices in $L^+(R, n)$ and hence $L(R, n)/R^* = L^+(R, n)/R^*_+$, $R^*_+ = R^* \cap L^+(R, n)$. In this case $L^+(R, n)/R^*_+ \cong L_1(R, n) \cong \mathfrak{J}$.

2. $L'(C, n)$. Similar considerations show that the automorphisms here are $a \to s^{-1}as$, $a \to s^{-1}(a')^{-1}s$, $a \to s^{-1}\bar{a}s$, $a \to s^{-1}(\bar{a}')^{-1}s$, $\bar{a} = (\bar{\alpha}_{ij})$ if $a = (\alpha_{ij})$. The first set has index 2 or 4 in $\mathfrak{A}$ according as $n = 2$ or $n > 2$ and is isomorphic to $L(C, n)/C^*$, C^* the set $\alpha 1$, $\alpha \neq 0$ in C. $L(C, n)/C^* \cong L'(C, n)/D \cong \mathfrak{J}$, D the set $\zeta 1$, where $\zeta^n = 1$.

3. $L'(Q, n)$. The automorphism are $a \to s^{-1}as$, $a \to s^{-1}(\bar{a}')^{-1}s$. If $n = 1$ the last set is superfluous and if $n > 1$ the first set has index 2 and is isomorphic to $L(Q, n)/R^* \cong L'(Q, n)/D \cong \mathfrak{J}$, $D = (1, -1)$.

5. Suppose S is an involutorial anti-automorphism in the associative algebra $\mathfrak{R}$, and $\mathfrak{U}'_S$ be the set of elements invariant under this automorphism $a \to (a^S)^{-1}$ of $\mathfrak{U}'$. Because of the continuity $\mathfrak{U}'_S$ is a closed subgroup of $\mathfrak{U}'$. If $a = \exp x$, $(a^S)^{-1} = \exp(-x^S)$ and $|x|$ is sufficiently small, $a = (a^S)^{-1}$ implies $x^S = -x$. Conversely if $x^S = -x$, x in R' then $\exp x$ is in $\mathfrak{U}'_S$. Thus the Lie algebra of $\mathfrak{U}'_S$ is $\mathfrak{S}'_S$ the derived algebra of $\mathfrak{S}_S$ the set of S-skew elements. Except for certain small values of n (cf. below) the algebras $\mathfrak{S}'_S$ are all simple and hence the $\mathfrak{U}'_S$ are t-simple. If S_1 and S_2 are cogredient anti-automorphisms in the same $\mathfrak{R}$ i.e. $S_2 = G^{-1}S_1G$, G an automorphism, it is evident that $\mathfrak{U}'_{S_1}$ and $\mathfrak{U}'_{S_2}$ are equivalent. Otherwise for n sufficiently large $\mathfrak{U}'_{S_1}$ and $\mathfrak{U}'_{S_2}$ are not locally isomorphic. If we identify $\mathfrak{U}'_{S_1}$ and $\mathfrak{U}'_{S_2}$ we find that the automorphisms of this group have the form $a \to a^G$ where G is an automorphism of $\mathfrak{R}$ commutative with S. Applying these principles we obtain the following t-simple groups and their automorphisms.

1. The real orthogonal groups $O(R, n, \nu)$ consisting of the real matrices a such that $a's_\nu a = s_\nu$, $\det a = 1$ where $s_\nu = e_{11} + \cdots + e_{pp} - e_{p+1,p+1} - \cdots - e_{nn}$, $\nu = 2p - n = n, n-2, \cdots, 0$ or 1 as n is even or odd and $n > 8$ if even, > 5 if odd. The group of automorphisms $\mathfrak{A}$ consists of the mappings $a \to b^{-1}ab$ where b satisfies $b's_\nu b = \rho s_\nu$, $\rho \neq 0$. Hence $\mathfrak{A}$ is the factor group of these matrices with respect to R^*. If n is odd $\rho > 0$ since the signatures of $b's_\nu b$ and s_ν are the same. We may choose in each class mod R^* a unique b_1 such that $b_1's_\nu b_1 = s_\nu$ and $\det b_1 = 1$. Thus $\mathfrak{A} = \mathfrak{J} \cong O(R, n, \nu)$. If n is even and $\nu \neq 0$, $\rho > 0$ and we may again choose a b_1 such that $b_1's_\nu b_1 = s_\nu$. However the elements of determinant 1 and -1 do not belong to the same class. The former form a subgroup of index $2 \cong O(R, n, \nu)/D \cong \mathfrak{J}$, $D = (1, -1)$. If n is even and $\nu = 0$ the representatives may be selected so that either $b_1's_0b_1 = s_0$ or $b_1's_0b_1 =$

$-s_0$. The former elements having det $= 1$ form an invariant subgroup of index $4 \cong O(R, n, \nu)/D, D = (1, -1)$.

2. The real symplectic group $S(R, 2n)$: a in R_n such that $a'qa = q, q' = -q$ and $n > 2$.[24] $\mathfrak{A}$ consists of $a \to b^{-1}ab$, $b'qb = \rho q$, $\rho \in R^*$. These elements for which $\rho > 0$ form a subgroup of index $2 = \mathfrak{J} \cong S(R, n)/D, D = (1, -1)$.

3. The complex orthogonal group $O(C, n)$: $a'a = 1$, a in $L'(C, n)$ and $n > 8$ if even, > 5 if odd. $\mathfrak{A}$ consists of $a \to b^{-1}ab$ and $a \to b^{-1}\bar{a}b$ where $b'b = \rho^1$, $\rho \in C^*$. The former form a subgroup of index $2 = \mathfrak{J} \cong O(C, n)/D, D = 1$, or $= (1, -1)$ according as n is odd or even.

4. The complex symplectic group $S(C, 2n)$: $a'qa = q$ $q' = -q$, a in C_n and $n > 2$.[24] $\mathfrak{A}$ consists of $a \to b^{-1}ab$ and $a \to b^{-1}\bar{a}b$, $b'qb = \rho q$, $\rho \in C^*$. $\mathfrak{A} = \mathfrak{J} \cong S(C, 2n)/D, D = (1, -1)$.

5. The complex unitary, unimodular groups $U'(C, n, \nu)$: $\bar{a}'s_\nu a = s_\nu$, s_ν and ν as in 1, $a \in L'(C, n)$, $n > 2$. $\mathfrak{A}$ consists of $a \to b^{-1}ab$ and $a \to b^{-1}\bar{a}b$, $\bar{b}'s_\nu b = \rho s_\nu$, $\rho \in R^*$. The former forms a subgroup of index 2 and unless n is even and $\nu = 0$, this subgroup $= \mathfrak{J} \cong U'(C, n, \nu)/D$, D the set $\zeta 1$, $\zeta^n = 1$. If n is even and $\nu = 0$, $\mathfrak{J}$ has index 4 in $\mathfrak{A}$.

6. The quaternionic unitary groups $U(Q, n, \nu)$; a in Q_n such that $\bar{a}'s_\nu a = s_\nu$, s_ν and ν as in 1 $n > 1$.[24] $\mathfrak{A}$ consists of $a \to b^{-1}ab$, $\bar{b}'s_\nu b = \rho s_\nu$, $\rho \in R^*$. As before $\mathfrak{A} = \mathfrak{J} \cong U(Q, n, \nu)/D, D = (1, -1)$ unless n is even, $\nu = 0$ when $\mathfrak{J}$ has index 2 in $\mathfrak{A}$.

7. The quaternionic symplectic group $S(Q, n)$: a in Q_n such that $\bar{a}'qa = q$ where $\bar{q}' = -q$, $n > 4$.[24] $\mathfrak{A}$ consists of $a \to b^{-1}ab$, $\bar{b}'qb = \rho q$, $\rho \in R^*$. The elements with $\rho > 0$ form a subgroup of index $2 = \mathfrak{J} \cong S(Q, n)/D, D = (1, -1)$.

The groups enumerated here are not locally isomorphic to any of the unimodular groups since their Lie algebras are not isomorphic. Altogether these ten classes include all t-simple Lie groups (relative to local isomorphism) except those of dimension 14, 28, 52, 78, 133, 248, 56, 104, 156, 266 and 498. The groups $O(R, n, n)$, $U'(C, n, n)$ and $U(Q, n, n)$ in our list are compact. This may be seen by noting that the conditions imposed entail the boundedness of the coördinates of these matrices.

8. Suppose F is any quasi-field, F_n the ring of $n \times n$ matrices over F and $L(F, n)$ the group of units in F_n. The method given by Dickson[25] for the case F a finite field may be applied to prove that any element of $L(F, n)$ has the form bd_n where b is a product of matrices of the type $1 + \theta e_{rs}$, $r \neq s$ and $d_n = 1 + (\delta - 1)e_{nn}$, $\delta \neq 0$.

We have

$$(1 + \theta e_{rs})(1 + (\delta - 1)e_{rr})(1 + \theta e_{rs})^{-1}(1 + (\delta - 1)e_{rr})^{-1} = 1 + (1 - \delta)\theta e_{rs}.$$

It follows that if F has more than two elements, b is a product of commutators.

---

[24] These elements all have determinant 1. Cf. v. d. Waerden 1, p. 10 for the cases $S(R, 2n)$ and $S(C, 2n)$. For the other cases see §8.

[25] Dickson [1], p. 78.

d_n is a commutator if δ is a commutator in F. If F is commutative and det $a = 1$, det $d_n = \delta = 1$ and hence a is a product of commutators.

This result applied to $F = R, = C$ shows that $L'(R, n) = L_1(R, n)$, $L'(C, n) = L_1(C, n)$ and $L^+(R, n)$ consists of the elements bd_n with $\delta > 0$. If $F = Q$ we employ the representation by complex matrices and obtain that $L'(Q, n) \leqq L_1(Q, n)$. Since det $b = 1$, det $d_n = N(\delta) = \delta\bar{\delta} > 0$. $N(\delta) = 1$ for a in $L_1(Q, n)$. If $\delta = 1$ it is evidently a commutator and if $\delta = -1$, $iji^{-1}j^{-1} = \delta$ ($1, i, j, k$ the quaternion units). If $\delta \neq 1, -1$, $R(\delta)$ is isomorphic to C and hence we have ξ in $R(\delta)$ so that $\xi^2 = \delta$, $\xi\bar{\xi} = 1$. These exists a μ such that $\mu\alpha = \bar{\alpha}\mu$ if $\alpha \epsilon R(\delta)$. Hence $\delta = \xi\mu\xi^{-1}\mu^{-1}$ and in all cases δ is a commutator. It follows that $L'(Q, n) = L_1(Q, n)$.

Any element θ in R, C or Q can be joined to 0 by an arc in these fields. It follows that the matrices b can be joined by arcs in $L'(R, n)$, $L'(C, n)$, $L'(Q, n)$ to 1. If δ is any element $\neq 0$ in C or Q or $\delta > 0$ in R then it can be joined to 1 by an arc which avoids the point 0. Finally if $N(\delta) = 1$ in Q, may be joined by an arc consisting of points of norm 1 to 1. It follows that $L^+(R, n)$, $L(C, n)$, $L(Q, n)$, $L'(R, n)$, $L'(C, n)$ and $L'(Q, n)$ are connected groups.

BIBLIOGRAPHY

G. Birkhoff

[1] "Analytical groups," *Trans. Am. Math. Soc.* 43 (1938), pp. 61–101.

E. Cartan

[1] *Thèse*, Paris, Nony 1894.

[2] "Les groupes réels simples et continus," *Annales de l'École Normale*, 31 (1914), pp. 263–355.

[3] "La théorie des groupes finis et continus et l'analysis situs," *Mem. des Sci. Math.* 1930.

[4] "Groupes simples clos et ouverts et géometrie riemanniene," *Jour. Math.* 2 (1929), pp. 1–33.

M. Deuring

[1] *Algebren, Ergebnisse der Math.*, 1935.

L. E. Dickson

[1] *Linear groups.*

L. P. Eisenhart

[1] *Continuous groups of transformations*, Princeton, 1933.

N. Jacobson

[1] "Rational methods in the theory of Lie algebras," *Annals of Math.* 36 (1935), pp. 875–881.

[2] "Simple Lie algebras over a field of characteristic zero," *Duke Math. Jour.* 4 (1938), pp. 534–551.

J. von Neumann

[1] "Gruppen von linearen Transformationen," *Math. Zeits.* 30 (1929), pp. 3–42.

O. Schreier

[1] "Abstrakte kontinuierliche Gruppen," *Hamb. Abhand.* 4 (1926), pp. 15–32.

B. L. van der Waerden

[1] *Gruppen von linearen Transformationen, Ergebnisse der Math.*, 1935.

University of North Carolina.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

ON A THEOREM OF MARSHALL HALL

BY WILHELM MAGNUS

(Received October 1, 1938)

It is the purpose of this note, to show that a simple proof of the theorem 4.1, as stated by Marshall Hall in his paper on "Group-rings and Extensions[1] I" can be given with the aid of a lemma, proved below, which also might be of some interest in itself.

Let us denote by $x, y, \cdots$ the generators of a group H, and by $\bar{\bar{x}}, \bar{\bar{y}}, \cdots$, a set of free generators of a free group F. By the correspondence $\bar{\bar{x}} \to x, \bar{\bar{y}} \to y, \cdots$, the free group F is mapped homomorphically onto H; therefore we have $F/R \simeq H$, where R is a self-conjugate subgroup of F. Let R' be the commutator—subgroup of R, and let us denote R/R' by A. Then the group $\bar{G} = F/R'$ is an extension of the abelian group A by the group H. Let $\bar{x}, \bar{y}, \cdots$ be the generators of $\bar{G}$, corresponding to the generators $\bar{\bar{x}}, \bar{\bar{y}}, \cdots$ of F. Hence we have $F \to \bar{G} \to H$, $\bar{\bar{x}} \to \bar{x} \to x$. The group A depends on the number n of generators of H; this number will not be the same throughout this paper, and we shall write $A(x, y, \cdots)$ instead of A, whenever this will be necessary to avoid confusion. The group A is the direct product of cyclic groups of infinite order. The number of the direct factors, or the rank of A, is equal to[2] $1 + (n - 1)j$, if j is the order of H.

Now there holds the following

LEMMA. *Let t_x, t_y, $\cdots$ be independent parameters, corresponding to $x, y, \cdots$ respectively, and permutable with all the elements of H. Then a true (that is, a one-to- one-isomorphic) representation of $\bar{G}$ by matrices is given by putting*

$$\bar{x} \to \begin{pmatrix} x & 0 \\ t_x & 1 \end{pmatrix}, \quad \bar{y} \to \begin{pmatrix} y & 0 \\ t_y & 1 \end{pmatrix}, \cdots. \tag{1}$$

We shall prove this lemma only in the case that H is a group of finite order j. If H is abelian, the lemma has been proved before.[3]

It is clear that, by the correspondence (1), a representation of $\bar{G}$ is given. For if $\phi(x, y, \cdots) = 1$ is a relation holding for the generators $x, y, \cdots$ of H, we have

$$\phi(\bar{x}, \bar{y}, \ldots) \to \begin{pmatrix} 1 & 0 \\ L & 1 \end{pmatrix}, \tag{2}$$

[1] Annals of Math. (2), **39**, pp. 220–234, 1938.

[2] s. O. Schreier, Abhandl. Math. Sem. der Hamburgischen Universität. **5**, p. 179, 1927.

[3] s. W. Magnus, *Über den Beweis des Hauptidealsatzes*, Journal für die reine und angewandte Mathematik **170**, pp. 235–240, 1934.

where L is a linear function of the parameters $t_x, t_y, \cdots$

$$L = t_x h_x + t_y h_y + \cdots, \tag{3}$$

the coefficients $h_x, h_y, \cdots$ being elements of the group-ring H^* of H. Therefore the group generated by the matrices (1) has a quotient-group isomorphic to H, the self-conjugate subgroup Λ corresponding to the identity of H being the additive group of certain linear forms of type (3). It is clear that the representation of $\bar{G}$ by (1) is a true one if and only if the number of the linearly independent forms (3) equals the rank of $A(x, y, \cdots)$, for Λ necessarily is a factor-group of A. This may be seen as follows. By definition, the group $\bar{G}$ may be considered as the "most general" group with generators $\bar{x}, \bar{y}, \cdots$, having the following property: Whenever $\varphi_i(x, y, \cdots) = 1$, $\varphi_j(x, y, \cdots) = 1$ in H, then the elements $\varphi_i(\bar{x}, \bar{y}, \cdots)$ and $\varphi_j(\bar{x}, \bar{y}, \cdots)$ of $\bar{G}$ are permutable; here the expression "most general" means that every group with generators $\bar{x}, \bar{y}, \cdots$, having this property, is a quotient group of $\bar{G}$. Now the group generated by the matrices occurring in formula (1) actually has this property.

For later purposes, we may notice here the relations (using now the equality-sign instead of the arrow):

$$\bar{x}^{-1} = \begin{pmatrix} x^{-1} & 0 \\ -t_x x^{-1} & 1 \end{pmatrix}; \quad \bar{x}^{-1}\bar{y}\bar{x} = \begin{pmatrix} x^{-1}yx & 0 \\ -t_x x^{-1}yx + t_y x + t_x & 1 \end{pmatrix}. \tag{4}$$

$$\phi^x \equiv \bar{x}^{-1}\phi(\bar{x}, \bar{y}, \cdots)\bar{x} = \begin{pmatrix} 1 & 0 \\ Lx & 1 \end{pmatrix}. \tag{4'}$$

First, we shall prove the lemma in the special case that all the elements $u, v, \cdots$ of H, except the identity, are generators of H. In this case we have

$$\bar{u} = \begin{pmatrix} u & 0 \\ t_u & 1 \end{pmatrix}, \quad \bar{v} = \begin{pmatrix} v & 0 \\ t_v & 1 \end{pmatrix}, \cdots, \tag{5}$$

and, of course, $t_u \equiv 0$ if $u = 1$. By a theorem of Reidemeister,[4] $A(u, v, \cdots)$ is generated by the elements

$$\overline{uv}^{-1}\bar{u}\bar{v} = \begin{pmatrix} 1 & 0 \\ -t_{uv} + t_u v + t_v & 1 \end{pmatrix}, \tag{6}$$

and therefore we must prove that there are

$$1 + (j-2)j = (j-1)^2 \tag{7}$$

linearly independent forms among the functions

$$-t_{uv} + t_u v + t_v, \tag{8}$$

$(j-1)^2$ being the rank of $A(u, v, \cdots)$. Apparently, the $(j-1)^2$ products $t_u v (u \neq 1, v \neq 1)$ are linearly independent, and therefore the same is true for the $(j-1)^2$ functions (8) containing those products.

---

[4] Abhandl. Math. Sem. Hamburgischen Universität 5, p. 7, 1927.

Now we turn to the investigation of the general case. Suppose the lemma had been proved already if H has exactly $k + 1$ generators $x_1, \cdots, x_k, y$, and let us assume that the generator y can be eliminated by a relation

$$\rho(x_1, \cdots, x_k)y = 1. \tag{9}$$

Then we have to show: If the group Λ as defined after (3) has the rank $1 + kj$ for the group generated by the matrices

$$\bar{x}_i = \begin{pmatrix} x_i & 0 \\ t_{x_i} & 1 \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} y & 0 \\ t_y & 1 \end{pmatrix}, \quad (i = 1, \cdots, k), \tag{10}$$

then the corresponding group Λ^* defined by the matrices

$$\bar{x}_i = \begin{pmatrix} x_i & 0 \\ t_{x_i} & 1 \end{pmatrix}, \quad (i = 1, \cdots, k), \tag{11}$$

has at least the rank $1 + (k - 1)j$. Now we have

$$\rho(\bar{x}_1, \cdots, \bar{x}_k) = \begin{pmatrix} y^{-1} & 0 \\ L & 1 \end{pmatrix}, \tag{12}$$

where L is a linear function of $t_{x_1}, \cdots, t_{x_k}$ of the type occurring in (3). Therefore

$$\rho(\bar{x}_1, \cdots, \bar{x}_k)\bar{y} = \begin{pmatrix} 1 & 0 \\ Ly + t_y & 1 \end{pmatrix}. \tag{13}$$

This shows that we may pass from Λ to Λ^* by postulating the relation

$$t_y = -L\rho^{-1}(x_1, \cdots, x_k) \tag{14}$$

for the parameters $t_{x_1}, \cdots, t_{x_k}, t_y$, the coefficients of this relation being elements of the group-ring H^*. Now there are at most j linearly independent linear forms to be combined from

$$t_y + L\rho^{-1}(x_1, \cdots, x_k) \tag{15}$$

by multiplication with elements of H^*, for H^* contains exactly j linearly independent elements. Therefore, by adding the equation (14), the rank of Λ will be diminished at most by j. This completes the proof of the lemma.—From (4′), we easily can deduce the following

COROLLARY: *Given any quotient-group $\bar{G}/C$ of $\bar{G}$, where C is contained in A, we can construct a true representation of $\bar{G}/C$ by matrices of type* (1), *by postulating the existence of certain linear relations for the parameters $t_x, t_y, \cdots$, the coefficients of these relations being elements of the group-ring H^*.*

In proving the theorem 4.1, page 225, of the paper by Marshall Hall, we shall adapt the notations used there.

Let $\phi_i(x, y, \cdots) = 1$ be a set of relations holding for the generators of H. From (1) and (2), we have

$$\phi_i(\bar{x}, \bar{y}, \ldots) = \begin{pmatrix} 1 & 0 \\ L_i & 1 \end{pmatrix}, \tag{16}$$

where $L_i = t_x x_i^* + t_y y_i^* + \cdots$, and $x_i^*, y_i^*, \cdots$ are elements of the group-ring H^* of H. If h_i is an arbitrary element of H^* we have by (4′):

$$\phi_i^{h_i} = \begin{pmatrix} 1 & 0 \\ L_i h_i & 1 \end{pmatrix}. \tag{17}$$

Now we define elements $\xi, \eta, \cdots$ by

$$\xi = \begin{pmatrix} 1 & 0 \\ t_\xi & 1 \end{pmatrix}, \qquad \eta = \begin{pmatrix} 1 & 0 \\ t_\eta & 1 \end{pmatrix}, \ldots, \tag{18}$$

$t_\xi, t_\eta, \cdots$ being parameters of the same kind as $t_x, t_y, \cdots$. If u is an element of H, we have

$$\xi^u = \bar{u}^{-1}\xi\bar{u} = \begin{pmatrix} 1 & 0 \\ t_\xi u & 1 \end{pmatrix} \tag{19}$$

and generally, if h is an element of H^*, we have

$$\xi^h = \begin{pmatrix} 1 & 0 \\ t_\xi h & 1 \end{pmatrix}. \tag{19′}$$

The group formed by the elements ξ^h apparently is "operator free" in the sense defined by Hall (cf. p. 223).

Now we have

$$\xi\bar{x} = \begin{pmatrix} x & 0 \\ t_\xi x + t_x & 1 \end{pmatrix}, \tag{20}$$

and therefore we will find $\phi_i(\xi\bar{x}, \eta\bar{y}, \cdots)$ by substituting $t_\xi x + t_x$ for t_x etc. in the linear form L_i occurring in the matrix

$$\phi_i = \begin{pmatrix} 1 & 0 \\ L_i & 1 \end{pmatrix},$$

that is:

$$\phi_i(\xi\bar{x}, \eta\bar{y}, \ldots) = \begin{pmatrix} 1 & 0 \\ (t_\xi x + t_x)x_i^* + (t_\eta y + t_y)y_i^* + \cdots & 1 \end{pmatrix}. \tag{21}$$

Therefore we have

$$\phi_i(\xi\bar{x}, \eta\bar{y}, \ldots) = \phi_i(\bar{x}, \bar{y}, \ldots)\xi^{xx_i^*}\eta^{yy_i^*}\cdots. \tag{22}$$

On the other hand, it follows from (16) that $\prod \phi_i^{h_i}(\bar{x}, \bar{y}, \cdots) = 1$ then and only then, if

$$\sum x_i^* h_i = \sum y_i^* h_i = \cdots = 0. \tag{23}$$

This is the same as

$$\sum xx_i^* h_i = \sum yy_i^* h_i = \cdots = 0. \tag{24}$$

From (22) it follows that $xx_i^* = x_i$, $yy_i^* = y_i$, $\cdots$ in the sense x_i, y_i, $\cdots$ are defined by Hall. Therefore $\sum x_i h_i = \sum y_i h_i = \cdots = 0$ are the necessary and sufficient conditions for $\prod \phi_i^{h_i}(\bar{x}, \bar{y}, \cdots) = 1$. This is the statement of the theorem 4.1.

FRANKFURT A. M., GERMANY.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

ADDITIVE SET FUNCTIONS ON GROUPS*

By S. Bochner

(Received April 1, 1939)

Introduction

All distinctive results of the Lebesgue theory of integration are based on the fact that Lebesgue measure is completely (that is countably) additive. Finitely additive Jordan volume is not sufficient. The completeness of L_p-spaces, the Riesz-Fischer theorem, the one-to-one relation between integrable point functions and absolutely continuous set functions—all these features of the Lebesgue theory are absent in the theory of Riemann integration.

The limitations of Jordan volume and Riemann integral are prevalent in the theory of almost periodic functions, whether these functions are given on the open line[1] or more generally on any group $\mathfrak{G}$.[2] The mean-value of almost periodic functions has the properties of a Riemann integral, and following Daniell[3] and v. Neumann[4] we shall show in Section I that it can be reduced to a Jordan volume. However, if the group $\mathfrak{G}$ is not compact, it will in general not be possible to imbed the Jordan volume in a Lebesgue measure, and for that reason theorems of the type of the Riesz-Fischer theorem have no obvious analogues. For functions on the open line, Besicovitch has constructed an elaborate solution of the Riesz-Fischer problem; however this solution cannot be readily extended to groups in general.

In the present paper we will show that some distinctive properties of Lebesgue integrals can be so formulated that their essential parts will remain valid for Jordan volume as well, and we will supplement the general statements by special results involving mean-values of almost periodic functions and their Jordan volume on groups.

The principal result of section II is a generalization to Jordan volume of the theory of Nikodym.[5] Nikodym has shown for arbitrary Lebesgue measure that an additive set function $F(E)$ which is absolutely continuous is the integral of an integrable point function $f(x)$,

---

* A summary of the results has appeared in Proc. Nat. Acad. 25 (1939), pp. 158–160.

[1] For a full account see Besicovitch [2].

[2] See von Neumann [11].

[3] In Daniell [6] a Lebesgue measure is constructed implicitly as the basis for any given Lebesgue integral.

[4] See von Neumann [10]. Here again the resulting measure has Lebesgue properties. However the construction is aimed at explicitly, and is closer to our own construction than the one in Daniell's paper.

[5] Compare Saks [15], Chapter I, 13.

$$F(E) = \int_E f(x)\, dv.$$

Any Lebesgue integrable function $f(x)$ is the limit, in the norm $\int | f(x) | \,dv$, of Riemann integrable functions $g(x)$ (for instance, of finitely-valued functions). Nikodym's abovementioned result is equivalent with the statement that in the Banach space of additive set functions $F(E)$ with the norm

$$F(E) = \sup \sum_{\nu=1}^{n} | F(E_\nu) |$$

(the sets $E_1, \cdots, E_n$ being disjoint) the integrals of finitely-valued point functions are everywhere dense. In this form, we shall see, the result also holds for any Jordan volume. Thus in case of Jordan volume, our Banach space of set functions takes the place of the L_1 space, and is isometric to it if the given volume is completely additive. Similar generalizations are available for L_p spaces, $p > 1$. Our proofs are new, even in the case of Lebesgue measure.

On the interval $0 \leqq x < 1$ the absolute continuity of a point function $F(x)$ of bounded variation can be defined in two different ways. The classical definition requires that for any finite number of non-overlapping intervals (a_ν, b_ν) the sum

$$\sum_{\nu=1}^{n} | F(b_\nu) - F(a_\nu) |$$

shall tend to 0 as $| b_\nu - a_\nu |$ tends to 0. Another definition which we suggest in the present paper is based on the fact that the functions of bounded variation represent the functionals on the Banach space C of all (periodic) continuous functions $\varphi(x)$ with the norm $\max | \varphi(x) |$. We call a function absolutely continuous if, uniformly in all continuous functions $\varphi(x)$ for which $| \varphi(x) | \leqq 1$, the integral $\left| \int_0^1 \varphi(x) dF(x) \right|$ tends to 0 as $\int_0^1 | \varphi(x) | \,dx$ tends to 0. In other words, a functional on C is called absolutely continuous if in the unit-sphere of C it is continuous relative to the norm of the space L_1. It is easy to see that in the given case the two definitions lead to the same absolutely continuous functions. In section III we will prove an analogue for general groups (the interval $0 \leqq x < 1$ being the group of real numbers mod 1).

A. Plessner[6] has proved that $F(x)$ is absolutely continuous if the (periodic) function

$$e(h) = \int_0^1 | d_x[F(x + h) - F(x)] |$$

is continuous at the origin and hence continuous throughout. We shall make the weaker assumption that $e(h)$ is almost periodic (no continuity being postulated) and we will prove the assertion under this assumption on general groups.

[6] See Plessner [13]. For generalizations of the theorem in other directions see, e.g. A. P. Morse [9].

Finally, section IV deals with a theorem of M. Riesz, and section V with functions on the open line $-\infty < x < \infty$.

I. Point Functions

1. Jordan fields

We consider an arbitrary point set $\mathfrak{G}$ whose points will be denoted by $x, y, \cdots$. A *Jordan field* is a class $\mathfrak{F}$ of subsets of $\mathfrak{G}$ whose general element will be denoted by E with the properties:

1) *$\mathfrak{F}$ contains the empty set; if $E \in \mathfrak{F}$, then $\mathfrak{G} - E \in \mathfrak{F}$;*
2) *if $E_1, E_2 \in \mathfrak{F}$ then $E_1 \cdot E_2 \in \mathfrak{F}$, $E_1 + E_2 \in \mathfrak{F}$*

and a numerical function on $\mathfrak{F}$ (Jordan volume) which will be denoted by vE or by $|E|$ with the properties:

3) $0 \leqq vE \leqq 1, |0| = 0, |\mathfrak{G}| = 1;$[7]
4) *if $E_1 \cdot E_2 = 0$, then $|E_1 + E_2| = |E_1| + |E_2|$.*

2. Partitions

If $E \in \mathfrak{F}$ then a partition of E—it will be denoted by $\delta = \delta(E) = (E_\nu)$—is a representation of E as a finite sum of mutually exclusive elements E_ν ($\nu = 1, \cdots, n$); the set E_ν will be called an element of δ. If the set which is to be partitioned is not specified, the partition refers to the whole set $\mathfrak{G}$. The partition δ will be called *regular* if $|E_\nu| > 0$ for each ν. If $\delta = (E_\nu)$ and $\delta' = (E'_\mu)$ are two partitions of the same element E then we put $\delta < \delta'$(δ' is greater than δ) if each E'_μ is contained in some E_ν and $\delta \neq \delta'$. A sequence of partitions $\{\delta_n\}$ is called monotone (and regular) if $\delta_n \leqq \delta_{n+1}$, $n = 1, 2, \cdots$ (and each δ_n is regular). A sequence $\{\delta_n\}$ is called a *closed sequence,* if it is monotone and if δ_n consists of n element $E_{n\nu}$, $\nu = 1, \cdots, n$. Obviously in a closed sequence the first partition δ_1 consists of one element and each δ_{n+1} arises from δ_n by subdividing just one element of δ_n into two and leaving the others unaltered. It is easy to see that any *monotone sequence can be made into a subsequence of some closed sequence.*

3. Riemann integrals

Let $f(x)$ be a bounded real function on $\mathfrak{G}$. Corresponding to any partition $\delta = (E_\nu)$ we form the sums

$$\bar{M}(\delta) = \sum_\nu \sup_{x \in E_\nu} f(x) \cdot |E_\nu|, \qquad \underline{M}(\delta) = \sum_\nu \inf_{x \in E_\nu} f(x) \cdot |E_\nu|$$

and their limits

$$\bar{M} = \lim_\delta \bar{M}(\delta), \qquad \underline{M} = \lim_\delta \underline{M}(\delta).$$

The limits are formed in accordance with the convention that $\lim_\delta A(\delta)$ exists and is equal l if corresponding to any $\epsilon > 0$ there exists a $\delta_0 = \delta(\epsilon)$ such that

[7] We impose the convenient restriction that the volume of the total set is finite. In our applications only this case will play a rôle.

for $\delta > \delta_0$, $|A(\delta) - l| < \epsilon$. We shall also admit $l = +\infty$, in which case $A(\delta) > 1/\epsilon$ for $\delta > \delta(\epsilon)$.

Obviously $\bar{M} \geqq \underline{M}$. The function $f(x)$ belongs to the class R (Riemann integrable functions) if $\bar{M} = \underline{M}$. The common value of the upper and lower integral will be denoted by

$$(1) \qquad M_x f(x) \quad \text{or} \quad \int_G f(x)\,dv \quad \text{or} \quad \int f(x)\,dv.$$

The extension to bounded complex-value functions is immediate. A real or complex function $f(x)$ belongs to R if and only if, for $\delta = (E_\nu)$

$$\lim_{\delta} \sum_{\nu} |f(x) - f(y)|_{x,y \in E_\nu} \cdot |E_\nu| = 0.$$

The class R contains the constants, and is closed with respect to addition, multiplication and formation of uniform limits. As a consequence, if $f(x)$, $g(x) \in R$ and if $\varphi(u, v)$ is uniformly continuous on the set of values of $f(x)$, $g(x)$ then $\varphi(f(x), g(x)) \in R$.

Also by definition of (1), the characteristic functions

$$\omega_E(x) = \begin{cases} 1, & x \in E \\ 0, & x \in \mathfrak{G} - E \end{cases}$$

belong to R and $\int \omega_E(x)\,dv = |E|$.

The subclass R_0 of R shall consist of functions $f(x)$ for which corresponding to any $\epsilon > 0$ there exists a partition (E_ν) such that the oscillation of $f(x)$ is $< \epsilon$ on each E_ν. The subclass R_0 has all closure properties of R and contains all characteristic functions $\omega_E(x)$ and their linear combinations.

We will also consider the norms

$$\|f(x)\|_p = \left(\int_G |f(x)|^p\,dv\right)^{1/p} \qquad p \geqq 1.$$

The space R if provided with this norm will be denoted by R_p. It has all properties of a Banach space except completeness. The completions will be discussed in later sections.

Let $f(x) \in R$, and $|f(x)| \leqq K$, and let $\delta = (E_\nu)$ be any partition of $\mathfrak{G}$. Putting

$$b_\nu = \sup_{x \in E_\nu} f(x), \qquad c_\nu = \inf_{x \in E_\nu} f(x),$$

$$(2) \qquad a_\nu = \frac{1}{|E_\nu|} \int_{E_\nu} f(x)\,dv,$$

we have $|b_\nu| \leqq K$, $|c_\nu| \leqq K$, $|a_\nu| \leqq K$. Each of the functions

$$\varphi(x) = \sum_\nu b_\nu \omega_{E_\nu}(x), \qquad \psi(x) = \sum_\nu c_\nu \omega_{E_\nu}(x),$$

$$(3) \qquad g(x) = \sum_\nu a_\nu \omega_{E_\nu}(x)$$

is a *step-function* having a constant value on each element of δ. Obviously

$$\varphi(x) \geqq f(x) \geqq \psi(x), \qquad \varphi(x) \geqq g(x) \geqq \psi(x)$$

and hence $| f(x) - g(x) | \leqq \varphi(x) - \psi(x)$. Now

$$\int | f(x) - g(x) |^p \, dv \leqq \int | \varphi - \psi |^p \, dv \leqq (2K)^{p-1} \int (\varphi - \psi) \, dv$$

$$= (2K)^{p-1}(\bar{M}(\delta) - \underline{M}(\delta))$$

and therefore we conclude: *The step functions are dense in R_p; especially each $f(x)$ from R_p can be approximated by functions of the type* (3), *with coefficients defined by* (2).

Given δ, we shall denote (3) by $f_\delta(x)$ and call it *the projection of $f(x)$ on δ.*

4. Modules of functions

We shall now invert the construction of Riemann integrable functions.

THEOREM 1. *Let C be a class of functions $\{f(x)\}$ on $\mathfrak{G}$ each defined everywhere and bounded, and let $M_x f(x)$ be a number which is defined for all $f(x)$ and let the following properties hold:*

1) *C contains the function $f(x) = 1$;*
2) *if $f_1 \epsilon C$, $f_2 \epsilon C$ and c_1 and c_2 are constants, then $f_1 f_2 \epsilon C$, $c_1 f_1 + c_2 f_2 \epsilon C$;*
3) *if $f \epsilon C$, then $\bar{f} \epsilon C$;*
4) *if $f \epsilon C$, and f is real, then $| f | \epsilon C$;*
5) *$M_x 1 = 1$;*
6) *$M_x(c_1 f_1 + c_2 f_2) = c_1 M_x f_1 + c_2 M_x f_2$;*
7) *$M_x f(x) \geqq 0$ if $f(x) \geqq 0$;*
8) *If f_n converges uniformly to f, and $f_n \epsilon C$, then $f \epsilon C$, and $M_x f_n$ converges to $M_x f$.*

Then, there exists a Jordan field $\mathfrak{F}$ such that C belongs to the class R_0 on that field and that

$$M_x f(x) = \int_{\mathfrak{G}} f(x) \, dv. \tag{4}$$

Also the class C is dense in each R_p, $p \geqq 1$.

Before giving the construction of $\mathfrak{F}$ we observe that property 8) will not be needed immediately but will be used only much later. Also if property 8) is not present to start with it can be easily secured by a closure process. On the other hand if property 8) is given, 4) is a consequence of the other properties since $| f |$ is a uniform limit of polynomials in f.

The construction of $\mathfrak{F}$ is as follows. Corresponding to any set A (its characteristic function will be denoted by $\omega_A(x)$) we define the outer and inner Jordan volumes

$$v^*A = \inf M_x f(x), \qquad f(x) \epsilon C, \qquad f(x) \geqq \omega_A(x),$$

$$v_*A = \sup M_x g(x), \qquad g(x) \epsilon C, \qquad g(x) \leqq \omega_A(x).$$

In these definitions we may restrict ourselves to functions $f(x)$ for which $f(x) \leqq 1$ (and to functions $g(x)$ for which $g(x) \geqq 0$). In fact, if $f(x) \in C$ and $f(x) \geqq \omega_A(x)$, then by our assumptions the function

$$\varphi(x) = \min (f(x), 1) = \frac{1 + f(x) - |1 - f(x)|}{2}$$

belongs again to C, is again $\geqq \omega_A(x)$, and $M_x\varphi(x) \leqq M_xf(x)$.

Obviously $v^*A \geqq v_*A$. The class $\mathfrak{F}$ shall consist of those sets $A = E$ for which $v^*A = v_*A$, and $vE = v^*E = v_*E$. Since $v^*0 = v_*0 = 0$ we have $|0| = 0$, similarly we find $|\mathfrak{G}| = 1$. It follows from assumptions 1), 2), 5) and 6) that $v^*(\mathfrak{G} - A) = 1 - v_*A$ and $v_*(\mathfrak{G} - A) = 1 - v^*A$, hence if $E \in \mathfrak{F}$ then $\mathfrak{G} - E \in \mathfrak{F}$. It also follows that $v^*(A_1 + A_2) \leqq v^*A_1 + v^*A_2$ and if $A_1 \cdot A_2 = 0$, $v_*(A_1 + A_2) \geqq v_*A_1 + v_*A_2$. Thus if $E_1, E_2 \in \mathfrak{F}$, then $E_1 + E_2 \in \mathfrak{F}$, and, if $E_1 \cdot E_2 = 0$, then $|E_1 + E_2| = |E_1| + |E_2|$.

We next want to show that $E_1, E_2 \in \mathfrak{F}$ implies $E_1 \cdot E_2 \in \mathfrak{F}$. Now $v^*E = v_*E$ if and only if corresponding to any $\eta > 0$ there exist functions $f(x), g(x) \in C$, with $1 \geqq f(x) \geqq \omega_E(x) \geqq g(x) \geqq 0$ such that $M_x(f(x) - g(x)) \leqq \eta$. Let $f_1(x)$, $g_1(x)$ and $f_2(x)$, $g_2(x)$ be such pairs of functions for the sets E_1, E_2 respectively. Then

$$1 \geqq f_1(x)f_2(x) \geqq \omega_{E_1 \cdot E_2}(x) \geqq g_1(x)g_2(x) \geqq 0.$$

f_1f_2, $g_1g_2 \in C$ and, since

$$0 \leqq f_1f_2 - g_1g_2 \leqq (f_1 - g_1)f_2 + g_1(f_2 - g_2) \leqq (f_1 - g_1) + (f_2 - g_2),$$

we obtain $M_x(f_1f_2 - g_1g_2) \leqq \eta + \eta = 2\eta$. This completes the proof of our first assertion that $\mathfrak{F}$ is a Jordan field.

We next consider a fixed real $f(x) \in C$. We introduce the point sets

$$A_a = E\{f(x) \geqq a\}, \qquad -\infty < a < \infty,$$

the functions

$$f_a(x) = \min \{f(x), a\}$$

which again belong to C, and the numbers

$$\varphi(a) = M_xf_a(x).$$

$\varphi(a)$ is monotonely non-decreasing, also $\varphi(a) = 0$ for $a \leqq a_0 = \inf_x f(x)$, and $= M_xf(x)$ for $a \geqq a_1 = \sup_x f(x)$. For $a < b$,

$$\omega_{A_b}(x) \leqq \frac{f_b(x) - f_a(x)}{b - a} \leqq \omega_{A_a}(x). \tag{5}$$

Therefore

$$v_*A_a \geqq M_x \frac{f_{a+h}(x) - f_a(x)}{h}, \qquad h > 0,$$

and hence

$$v_* A_a \geqq D^+ \varphi(a);$$

similarly

$$v^* A_a \leqq D^- \varphi(a).$$

If for a given value a, $D^+\varphi(a) = D^-\varphi(a)$, then $v_* A_a = v^* A_a$. Thus A_a belongs to $\mathfrak{F}$ for a dense set of numbers a. Given $\eta > 0$, $\eta < \frac{1}{3}$ we can select from the dense set a finite sequence $a = b_\nu$, $\nu = 0, 1, \cdots, n$ for which

$$a_0 - 1 = b_0 < b_1 < \cdots < b_n = a_1 + 1 \quad \text{and}$$

$b_{\nu+1} - b_\nu < \eta$. Putting $E_\nu = A_{b_{\nu-1}} - A_{b_\nu}$, we have

$$\sup_{x, y \in E_\nu} | f(x) - f(y) | \leqq (b_\nu - b_{\nu-1}) < \eta,$$

and this proves our assertion that $f(x) \in R_0$. Also

$$\sum_{\nu=1}^{n} b_{\nu-1} \, | E_\nu | \leqq \int_G f(x) \, dv \leqq \sum_{\nu=1}^{n} b_\nu \, | E_\nu |. \tag{6}$$

On the other hand,

$$\varphi(b_n) - \varphi(b_0) = \sum_{\nu=1}^{n} \frac{\varphi(b_\nu) - \varphi(b_{\nu-1})}{b_\nu - b_{\nu-1}} (b_\nu - b_{\nu-1})$$

and by (5),

$$| A_{b_\nu} | \leqq \frac{\varphi(b_\nu) - \varphi(b_{\nu-1})}{b_\nu - b_{\nu-1}} \leqq | A_{b_{\nu-1}} |$$

and hence

$$\sum_{\nu=1}^{n} b_{\nu-1} \, | E_\nu | + b_n A_{b_n} \leqq \varphi(b_n) + (b_0 A_{b_0} - \varphi(b_0)) \leqq \sum_{\nu=1}^{n} b_\nu \, | E_\nu | + b_n A_{b_n}.$$

Since $A_{b_0} = 1$, $\varphi(b_0) = b_0$, $A_{b_n} = 0$, we finally have

$$\sum_{\nu=1}^{n} b_{\nu-1} \, | E_\nu | \leqq \varphi(b_n) = M_x f(x) \leqq \sum_{\nu=1}^{n} b_\nu \, | E_\nu |.$$

Comparing this with (6) we obtain (4).

The last statement of theorem 1 will be satisfied if we show that for any $E \in \mathfrak{F}$, $\int | f(x) - \omega_E(x) |^p \, dv$ can be made arbitrarily small for an appropriate choice of $f(x) \in C$. Assuming $1 \geqq f(x) \geqq \omega_E(x)$ the integral is $\leqq \int (f(x) - \omega_E(x)) \, dv = M_x f(x) - | E |$ and this can indeed be made as small as we please. Any class of functions having the properties of theorem 1 will be called a *module,* and the Jordan field $\mathfrak{F}$ whose construction we have just described will be called a *generated Jordan field* and it will be denoted explicitly by $\mathfrak{F}_C$.

5. Lebesgue fields

A *Lebesgue field* is a Jordan field having the following properties: (i) *If* $E_\nu \in \mathfrak{F}$, $\nu = 1, 2, \cdots$, *then* $E_1 + E_2 + \cdots \in \mathfrak{F}$, (ii) *If* E_ν *converges monotonely to* E, *then* $| E_\nu |$ *converges to* $| E |$ and (iii) *Any subset of a set of volume* 0 *is again an element of the field.*

THEOREM 2. *A generated Jordan field* $\mathfrak{F}_C$ *can be extended to a Lebesgue field if and only if for any sequence* $\{f_n(x)\} \subset C$, *the assumptions*

$$\lim_{n \to \infty} f_n(x) = 0, \qquad |f_n(x)| \leqq K \tag{7}$$

imply

$$\lim_{n \to \infty} M_x f_n(x) = 0. \tag{8}$$

In fact if $\mathfrak{F}$ is part of a Lebesgue field then $M_x f(x)$ is a Lebesgue integral and (8) is a consequence of (7). Conversely it is known that $\mathfrak{F}$ can be extended to a Lebesgue field if the assumptions

$$E_1 \supset E_2 \supset E_3 \supset \cdots; \qquad \lim E_n = 0 \tag{9}$$

imply

$$\lim | E_n | = 0.\text{[8]} \tag{10}$$

Now corresponding to each E_n we pick a function $f_n(x) \in C$ such that $0 \leqq f_n(x) \leqq \omega_{E_n}(x)$ and $| E_n | - M_x f_n(x) \to 0$. Obviously (9) implies (7), this implies (8), and this again implies (10).

An an *illustration* consider the *module* C consisting of all continuous almost periodic functions on the line $-\infty < x < \infty$ with the mean value

$$M_x f(x) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, dt,$$

say. This module will be called the *module of Bohr functions*. The sequence $f_n(x) = \left(\sin \frac{x}{n}\right)^2$ converges to 0 at every point, and yet

$$M_x \left(\sin \frac{x}{n}\right)^2 = M_x \frac{1 - \cos \frac{2x}{n}}{2} = \frac{1}{2}.$$

Thus the Jordan field generated by C cannot be extended to a Lebesgue field.

Our module C can also be used to illustrate *another impossibility*. Let $g(x)$, $h(x) \in C$, and let $\omega(x)$ be any function on $-\infty < x < \infty$ which is Lebesque integrable in the ordinary sense, such that $g(x) \leqq \omega(x) \leqq h(x)$. Obviously

$$M_x g(x) \leqq \underline{\lim}\, \frac{1}{2T} \int_{-T}^{T} \omega(t)\, dt \leqq \overline{\lim}\, \frac{1}{2T} \int_{-T}^{T} \omega(t)\, dt \leqq M_x h(x). \tag{11}$$

[8] See for instance Jessen [7].

Now, if $f(x) \in C$, then for any a, the point set $A_a = E\{f(t) \geqq a\}$ is measurable in the ordinary Lebesgue measure. Hence, if for a special value of a, $A_a \in \mathfrak{F}$, then, by (11),

$$| A_a | = \lim \frac{1}{2T} \int_{-T}^{T} \omega_{A_a}(x)\, dx.$$

But it was shown by H. Bohr that the latter limit need not exist for every a if $f(x)$ is an arbitrary almost periodic function.[9] Therefore it was essential in our proof of theorem 1 to argue the existence of a dense set of numbers $\{a\}$, for which $E_a \in \mathfrak{F}$; the property does not necessarily hold for every a.

II. Spaces of Set Functions

6. Definitions

Apart from point functions we shall also consider set function. Every set function $F = F(E)$ will be defined on all sets E of some Jordan field $\mathfrak{F}$ and it will be finitely additive and of bounded variation. Point functions will be denoted by small letters, set functions by capital letters. If a set function is the indefinite integral of a point function, it will be denoted by the same letter. Thus if $f(x) \in R$, then its indefinite integral will be denoted by $F(E)$, and so we have

$$F(E) = \int_E f(x)\, dv.$$

Similarly the indefinite integral of f_δ, φ, g_n will be denoted by F_δ, Φ, G_n. Also we shall say that the set function $F = F(E)$ belongs to R, R_0, R_p, C etc. if it is the indefinite integral of a point function belonging to R, R_0, R_p, C etc. But we wish to stress that unless otherwise stated a symbol like $F(E)$ will denote a set function which need not be an indefinite integral.

If $| E_0 | > 0$, $F(E)$ is said to be constant on E_0, if for $E \subset E_0$

$$\frac{F(E)}{| E |} = \frac{F(E_0)}{| E_0 |} \tag{12}$$

(which implies that $F(E) = 0$ if $| E | = 0$). The common value of the fractions (12) will be called the "value" of $F(E)$ on E_0. If $\delta = (E_\nu)$ is a regular partition of $\mathfrak{G}$, then F is called a step function on δ if it is constant on each E_ν. Also if a_ν are its values on E_ν, then $F(E)$ is the indefinite integral of $\sum_\nu a_\nu \omega_{E_\nu}(x)$.

In particular if $F(E)$ is *any* set function and $\delta = (E_\nu)$ is any regular partition, then the step-function with the values $a_\nu = \dfrac{F(E_\nu)}{| E_\nu |}$ on δ will be called the *projection of* $F(E)$ *on* δ and will be denoted by $F_\delta(E)$. Projections will be an important means in our study of set functions.

---

[9] See Bohr [5].

7. Norms

Given $p \geqq 1$, we consider for any $F(E)$ and any $\delta = (E_\nu)$ the expression

$$A(\delta) = \left(\sum_\nu \frac{|F(E_\nu)|^p}{|E_\nu|^{p-1}}\right)^{1/p}. \tag{13}$$

If $p > 1$, in order to avoid unnecessary complications we shall form (13) only for such functions $F(E)$ for which $F(E_0) = 0$ if $|E_0| = 0$, and we agree that for $|E_\nu| = 0$ the term $|F(E_\nu)|^p/|E_\nu|^{p-1}$ in (13) shall have the value 0. $A(\delta)$ is monotone in δ that is $A(\delta) \leqq A(\delta')$ if $\delta < \delta'$; therefore it approaches a limit as δ increases; also the expression

$$\|F\| = \|F\|_p = \lim_\delta A(\delta)$$

has the properties of a Banach norm. Moreover by our definition of the projection F_δ, $A(\delta) = \|F_\delta\|$, and thus

$$\|F\| = \lim_\delta \|F_\delta\|.$$

The resulting complete[10] Banach space will be denoted by V_p, $p \geqq 1$. The space V_1 has a complicated structure; we will be interested also in its closed subspace AC (space of absolutely continuous functions). A function $F(E)$ belongs to AC if there exists a function $\eta(\epsilon)$, $\lim_{\epsilon\to0} \eta(\epsilon) = 0$, such that $|E| \leqq \eta(\epsilon)$ implies $|F(E)| \leqq \epsilon$. We further add the Banach space V_∞ with the norm

$$\|F\| = \|F\|_\infty = \sup_E \frac{|F(E)|}{|E|}, \qquad (|E| \neq 0).$$

As a rule we will assume $p \neq \infty$.

If $F \epsilon V_p$, $p > 1$, or $F \epsilon AC$, then there exists a monotone sequence $\{\delta_n\}$ such that

$$\|F\| = \lim_{n\to\infty} \|F_{\delta_n}\|. \tag{14}$$

Now, we may assume that the sequence is closed and regular. In fact let $\{\delta_n'\}$ be any monotone sequence for which (14) holds. Let E_0 be any element of δ_{n-1}' and let $|E_0| > 0$. There are elements of δ_n whose sum is E_0, we shall denote them by $E_1, \cdots, E_\lambda$. We assume that $|E_1| > 0, \cdots, |E_\mu| > 0$ and $|E_{\mu+1}| = \cdots = |E_\lambda| = 0$. Replacing in δ_n' the sets $E_1, \cdots, E_\lambda$ by the sets $E_1, E_2, \cdots, E_{\mu-1}, E_\mu + E_{\mu+1}, + \cdots + E_\lambda$ we will obviously not alter the value of $\|F_{\delta_n'}\|$. Carrying out this contraction for successive values of $n = 1, 2, \cdots$ and all sets E_0 we obtain a new monotone sequence $\{\delta_n''\}$. It is regular and it satisfies (14). Finally the latter sequence can be made the subsequence of a sequence $\{\delta_n\}$ which is also closed.

---

[10] It is not hard to conclude by the usual reasoning that any Cauchy sequence is convergent towards a limit element.

Obviously, V_p, $p > 1$, contains $\mathrm{Cl}(R_p)$, (closure of R_p), and AC contains $\mathrm{Cl}(R_1)$. We shall prove that these pairs of spaces are actually identical. A special proof will be used for the spaces V_p, although the method which will apply for the space AC could also be used in the other cases.

8. An inequality for $p > 1$

For $p > 1$ and $\epsilon > 0$ there exists a finite number $K(p, \epsilon)$ such that for all real numbers f,

$$(15) \qquad |f - 1|^p \leqq K(p, \epsilon)(|f|^p + p - 1 - pf) + \epsilon |f|^p.$$

In fact, the function $\lambda(f) = |f|^p + p - 1 - pf$[11] is continuous in $-\infty < f < \infty$ and > 0 for $f \neq 1$, and $|f - 1|^p = 0(\lambda(f))$ as $|f| \to \infty$. Thus $|f - 1|^p \leqq M(p, \eta)\cdot\lambda(f)$ for $|f - 1| > \eta$, $\eta > 0$. Now given any $\epsilon > 0$ there exists an $\eta > 0$, such that $|f - 1| \leqq \eta$ implies $|f - 1|^p \leqq \epsilon |f|^p$, and this proves (15).

THEOREM 3. *For $p > 1$ and $\epsilon > 0$ the inequality*

$$(16) \qquad \|f - f_\delta\|^p \leqq K(p, \epsilon)(\|f\|^p - \|f_\delta\|^p) + \epsilon \|f\|^p$$

holds for any $f(x) \in R$ and any partition $\delta = (E_\nu)$.

Let $f(x)$ be any real function from R, and let E be any set for which $\int_E f(x)\, dv = |E|$. Putting $f = f(x)$ in (15) and integrating over E we obtain

$$(17) \qquad \int_E |f(x) - 1|^p\, dv \leqq K(p, \epsilon)\left(\int_E |f(x)|^p\, dv - |E|\right) + \epsilon \int_E |f(x)|^p\, dv.$$

Hence we obtain for any $f(x)$ in R,

$$\int_E |f(x) - a|^p\, dv \leqq K(p, \epsilon)\left(\int_E |f(x)|^p\, dv - |a|^p |E|\right) + \epsilon \int_E |f(x)|^p\, dv,$$

where

$$a = \frac{1}{|E|}\int_E f(x)\, dv.$$

Putting $E = E_\nu$, and adding for all ν we obtain (16).

THEOREM 4. *For $p > 1$ and $\epsilon > 0$ the inequality*

$$(18) \qquad \|F - F_\delta\|^p \leqq K(p, \epsilon)(\|F\|^p - \|F_\delta\|^p) + \epsilon \|F\|^p$$

holds for every $F \in V_p$.

For every δ, F_δ is a step function in R_p, and for $\delta' > \delta$, $(F_{\delta'})_\delta = F_\delta$. Therefore, by theorem 3,

$$(19) \qquad \|F_{\delta'} - F_\delta\|^p \leqq K(p, \epsilon)(\|F_{\delta'}\|^p - \|F_\delta\|^p) + \epsilon \|F_{\delta'}\|^p$$

for $\delta' > \delta$. Taking the limit with respect to δ', $\|F_{\delta'}\|$ tends to $\|F\|$ and $\|F_{\delta'} - F_\delta\| = \|(F - F_\delta)_{\delta'}\|$ tends to $\|F - F_\delta\|$, and (18) follows from (19).

[11] This function has been introduced in F. Riesz [14].

9. The structure of V_p, $p > 1$

As an immediate consequence of theorem 4 we have

THEOREM 5. *If $F \in V_p$, $p > 1$, and $\{\delta_n\}$ is a sequence of partitions, then $\| F_{\delta_n} \| \to \| F \|$ implies $\| F - F_{\delta_n} \| \to 0$.*

Since each F_δ is a step function we hence obtain

THEOREM 6. *The step-functions are dense in V_p, $p > 1$. Therefore, $V_p = \mathrm{Cl}\,(R_p)$, $p > 1$.*

On the basis of theorems 6 and 5 various properties of ordinary L_p spaces can be established. We shall collect them into one theorem, and omit the proofs.

THEOREM 7. *The spaces V_p and V_q, $p > 1$, $q > 1$, $p = q/q - 1$, are conjugate. Corresponding to any functional $X(F)$ on V_p there exists an element $G \in V_q$, such that*

$$X(F) = \lim_{\delta} M_x(f_\delta(x) g_\delta(x))$$

and vice versa (the projections $F_\delta(E)$, $G_\delta(E)$ being the indefinite integrals of the point functions $f_\delta(x)$, $g_\delta(x)$), and the norm of $X(F)$ is $\| G \|_q$.

A sequence $\{F_\nu\}$, $F_\nu \in V_p$ is weakly convergent if $\| F_\nu \| \leqq K$ and

$$\lim F_\nu(E) \qquad \nu \to \infty$$

exists for each E, the limit defining the limit element. Also, if $\{F_\nu\}$ converges weakly to 0, then for a suitable subsequence F'_μ

$$\left\| \sum_{\mu=1}^{m} F'_\mu \right\|_p = \begin{cases} 0(m^{1/p}) & \text{if} \quad 1 < p \leqq 2 \\ 0(m^{1/2}) & \text{if} \quad p \geqq 2.^{12} \end{cases}$$

Finally, if V_p is separable it has a basis consisting of step functions.[13]

10. A mapping process[14]

Let $\{\delta_n\}$ be any closed regular sequence of partitions of $\mathfrak{S}$. Denoting the elements of δ_n by $E_{n\nu}$, $\nu = 1, \cdots, n$, we may assume that for each n there exists a μ such that $E_{n+1,\nu} = E_{n,\nu}$ for $\nu = 1, \cdots, \mu$, $E_{n+1,\mu+1} + E_{n+1,\mu+2} = E_{n,\mu+1}$ and $E_{n+1,\nu} = E_{n,\nu-1}$ for $\nu = \mu + 3, \cdots, n + 1$. We now form the numbers

$$(20) \qquad t_{no} = 0, \qquad t_{n\nu} = \sum_{\mu=1}^{\nu} | E_{n\mu} |, \qquad \nu = 1, \cdots, n, \qquad n = 1, 2, \cdots,$$

and with these numbers the linear intervals

$$t_{n,\nu-1} \leqq t < t_{n,\nu}$$

---

[12] The proof in Banach [1], p. 197–9 is immediately applicable to set functions which are integrals of point function and, by approximation, to set functions in general.

[13] The proof of this statement will be included in another publication.

[14] This process has been applied repeatedly to measure in product spaces.

which we denote by $e_{n\nu}$. Obviously for each n the set of intervals $(e_{n\nu})$, $\nu = 1, \cdots, n$, is a partition of the linear interval $0 \leqq t < 1$,—we will denote it by $\delta_n(t)$—, and the sequence $\{\delta_n(t)\}$ is a closed regular sequence of such partitions. Also

$$| E_{n\nu} | = | e_{n\nu} | = t_{n,\nu} - t_{n,\nu-1} .$$

The set of numbers (20) will be denoted by T_0, its compact closure by T. Any interval whose endpoints are points of T will be denoted by i (with a subscript). If T is not the whole interval $0 \leqq t \leqq 1$, let its complement consist of the (open) intervals $\{\bar{\imath}_\nu\}$, $\nu = 1, 2, \ldots$. Any interval which after omission of its left end point is a proper part of an interval $\bar{\imath}_\nu$ will be denoted by i^* (with a subscript). An arbitrary interval e of $0 \leqq t < 1$ is the sum of an interval i and an interval i^*.

Now let $F(E)$ be an arbitrary absolutely continuous function. Our mapping process gives rise to a function $F(E)$ on $0 \leqq t < 1$. The latter function is defined at first on the intervals whose end points are points of T_0. Since $F(E)$ is absolutely continuous, $F(e)$ varies continuously with the end points of e, and thus we can extend $F(e)$ onto all intervals i. In particular we can define it for all intervals $\{\bar{\imath}_\nu\}$. Finally we define it on the parts i^* of $\bar{\imath}_\nu$ by linear extension

$$F(i^*) = \frac{| i^* |}{| \bar{\imath}_\nu |} F(\bar{\imath}_\nu).$$

After additive extension the resulting function $F(E)$ exists on all intervals e of $0 \leqq t < 1$. Our final step is to show that $F(e)$ is again absolutely continuous. We have to prove that existence of a function $\eta(\epsilon)$, $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$, such that

$$\sum_{\kappa=1}^{k} | F(e_\kappa) | < \eta\left(\sum_{\kappa=1}^{k} | e_\kappa |\right) \tag{21}$$

on the assumption that there exists a function $\eta_0(\epsilon)$, $\lim_{\delta \to 0} \eta_0(\epsilon) = 0$ for which

$$\sum_{\lambda=1} | F(E_\lambda) | < \eta_0\left(\sum_{\lambda=1} | E_\lambda |\right); \tag{22}$$

the sets (e_κ) and the sets (E_λ) being disjoint. Since subdivision of the intervals e_κ does not decrease the left side of (21), we may assume for the sake of the proof that either 1) all e_κ are intervals i_κ, or 2) all e_κ are intervals i_κ^*. In the first case we may obviously put $\eta(\epsilon) = \eta_0(\epsilon)$. In the second case we choose n so large that $\sum_{n+1}^{\infty} | \bar{\imath}_\nu | < \epsilon/2$. For those i^* which lie in $\sum_{n+1}^{\infty} \bar{\imath}_\nu$ we have $\sum | F(i_\kappa^*) | \leqq \sum_{n+1}^{\infty} | F(\bar{\imath}_\nu) | < \eta_0(\epsilon/2)$. The number n being fixed and $F(e)$ being constant on each $\bar{\imath}_\nu$ we can obviously find an $\eta_1(\epsilon/2)$ such that for $\sum_\kappa i_\kappa^* \subset \sum_1^n \bar{\imath}_\nu$

$$\sum_\kappa | F(i_\kappa^*) | < \eta_1\left(\sum_\kappa | i_\kappa^* |\right).$$

This completes the proof of our statement.

11. The structure of AC

THEOREM 8. *If $F(E) \epsilon AC$, and if $\{\delta_n\}$ is a closed regular sequence of partitions, then*

$$\lim \| F_{\delta_m} - F_{\delta_n} \| = 0, \qquad m, n \to \infty.$$

The proof will require several steps.

We shall temporarily call a function $G(E)$ a step function on $\{\delta_n\}$ if it is a step function on some δ_ν (ν sufficiently large). Our theorem is trivially true for such functions. Therefore it is also true for functions which are limits in the norm of such functions, since,

$$\| F_{\delta_m} - F_{\delta_n} \| \leqq \| (F - G)_{\delta_m} \| + \| G_{\delta_m} - G_{\delta_n} \| + \| (F - G)_{\delta_n} \|$$
$$\leqq 2 \| F - G \| + \| G_{\delta_m} - G_{\delta_n} \|.$$

We next apply our mapping process, as described in §10, relative to the given sequence $\{\delta_n\}$ and we thus obtain, on the interval $0 \leqq t < 1$, a closed sequence $\{\delta_n\} = \{\delta_n(t)\}$ and a function $F(e)$. Clearly, $\| F(e) \| \leqq \| F(E) \|$, the equality *not* holding in general. The equality does hold, if $F(E)$ is a step function on $\{\delta_n\}$, and therefore

$$\| F_{\delta_m}(E) - F_{\delta_n}(E) \| = \| F_{\delta_m}(e) - F_{\delta_n}(e) \|.$$

Combining this with our first step, we see that our theorem will be proved if we show that our function $F(e)$ is the limit in norm of functions $G(e)$ which are step functions on $\{\delta_n(t)\}$.

Now by the theory of Lebesgue, since $F(e)$ is absolutely continuous there exists, for given $\epsilon > 0$, some step function $G(e)$ for which

$$\| F(e) - G(e) \| < \epsilon. \tag{23}$$

If $\delta(t) = (e_\kappa)$ is the partition on which $G(e)$ is a step function, we may assume that any e_κ is either an interval i or an interval i^*. If e_κ is an interval i^* lying in an interval $\bar{\imath}_\nu$, then there must be other intervals e_ν whose sum is the given $\bar{\imath}_\nu$. Now $F(e)$ is constant on $\bar{\imath}_\nu$. Hence replacing $G(e)$ on $\bar{\imath}_\nu$ by the constant value of $F(e)$ will lead to a new step function $G(e)$ for which (23) holds too. Carrying out this adjustment for all intervals i^*, we will obtain a step function $G(e)$ for which the intervals of constancy are intervals i. Slight variations of the intervals will make them into intervals whose end points are points of T_0. Thus (23) can be satisfied by a function $G(e)$ which is a step function on $\delta_n(t)$.

This completes the proof of our theorem.

THEOREM 9. *The step functions are dense in AC, thus $AC = \mathrm{Cl}\,(R_1)$.*

Suppose the theorem were false, and let $F(E)$ be a function which cannot be approximated by step functions. Then there exists a positive constant λ such that for each regular partition δ, $\| F - F_\delta \| > \lambda$. Since

$$\lim_{\delta'} \| F_{\delta'} - F_\delta \| = \lim_{\delta'} \| (F - F_\delta)_{\delta'} \| = \| F - F_\delta \|,$$

if δ is given, we can find a regular partition $\delta' > \delta$ such that $\| F_{\delta'} - F_\delta \| > \lambda$. In this way we could construct a closed regular sequence for which theorem 8 does not hold.

As a consequence of these theorems the following theorem could be proved

THEOREM 10. *The conjugate space to AC is V_∞. If $G \epsilon V_\infty$, the functional it represents can be written in the form*

$$X(F) = \lim_\delta \int f_\delta(x)\, dG.$$

A sequence $\{F_\nu\}$ is weakly convergent in AC if the norms $\| F_\nu \|$ are bounded and the functions $\{F_\nu\}$ are uniformly absolutely continuous, and if the limit of $F_\nu(E)$ exists for each E.

If AC is separable it has a basis consisting of step function.

12. Functions of bounded variation on generated fields

Let $\mathfrak{F}$ be a Jordan field which is generated by a module C. The norm

$$\| f \| = \sup_x | f(x) |$$

makes C into a Banach space, its conjugate space will be denoted by V_C. The latter space arises from V_1 by identification of some of the elements and contraction of the norm. In fact, if $G \epsilon V_1$, then the expression

$$X(f) = \int f(x)\, dG \tag{24}$$

represents a functional on C, that is an element of V_C, and its norm is

$$\| G \|_C = \sup_f \left| \int f(x)\, dG \right| \qquad \| f \| \leqq 1.$$

Obviously, $\| G \|_C \leqq \| G \|_1$. Conversely, any element of V_C can be so represented by at least one element G of V_1. In order to find such an element we extend the functional, without increasing its norm, from the given space C to the larger space R_0, and in the latter space every functional can be represented by a Stieltjes integral (24) the function $G(E)$ being defined by

$$G(E) = X(\omega_E).$$

In order to avoid the introduction of new symbols we shall denote the elements of V_C in the same way as the element of V_1. Elements of V_1 will be called equivalent if they represent the same element of AC; equivalence will be denoted by "$\simeq$."

The element $I(E) = | E |$ of V_1 or its equivalents will be called the *identity* in V_C.

The subspace of V_C which corresponds to the subspace AC of V_1—we shall denote the new subspace by V_{AC}—is the closure of the set of elements $G \epsilon V_C$ which are indefinite integrals of elements $g \epsilon C$. It can be shown that an ele-

ment G of V_C belongs to V_{AC} if corresponding to any $\epsilon > 0$ there exists a $\eta(\epsilon)$, $\lim_{\epsilon\to 0} \eta(\epsilon) = 0$, such that for $|f(x)| \leqq 1$,

$$\left| \int f(x)\, dG \right| \leqq \eta\left(\int |f(x)|\, dv \right).$$

More remarkable properties of the spaces V_C and V_{AC} will be established for modules on groups.

III. Set Functions on Groups

13. Group invariant modules

If $\mathfrak{G}$ is a group, and if C consists of almost periodic functions, it is appropriate to assume that C is group invariant; that is if $f(x)$ belongs to C then $f(x^{-1})$ belongs to C, and if a is an arbitrary element of G, then the functions $g(x) = f(xa)$, $h(x) = f(ax)$ again belong to C. The generated Jordan field is also group invariant; if E belongs to $\mathfrak{F}$ then the sets E^{-1}, Ea (right-translation of E by a) and aE (left translation of E by a) also belong to F, and $|E^{-1}| = |Ea| = |aE| = |E|$.

By our definition of equivalence, two elements $F \epsilon V_1$ are equivalent, if the integrals $\int \varphi(x)\, dF$ have the same value for all $\varphi \epsilon C$. In particular, $F \simeq 0$, if these integrals are all 0. For general C, it would not be easy to decide, whether "equivalence" implies identical equality, even if we assume that the elements F belong to AC. But for group invariant modules this is always the case.

Theorem 11. *If C is group invariant, if $F \epsilon AC$ and if $\int \varphi(x)\, dF = 0$ for all $\varphi \epsilon C$, then $F(E) = 0$.*

The proof will require several steps.

1) If $f(x) \epsilon C$, then corresponding to any $\epsilon > 0$ there exists another function $g(x) \epsilon C$ with the following properties: (i) $g(x) \geqq 0$, (ii) $M_x g(x) = 1$, and (iii) putting $h(x) = M_y g(xy^{-1}) f(y)$, then

$$\|f - h\|_1 \leqq \epsilon. \tag{25}$$

In fact the function $g(x)$ can be chosen to satisfy the stronger relation $|f - h| \leqq \epsilon$.[15]

2) If $F(E) \epsilon AC$ and $g(x) \epsilon C$, the function

$$h(x) = \int g(xy^{-1})\, d_y F \tag{26}$$

again belongs to C, and

$$\|h\|_1 \leqq (M_x |g(x)|) \cdot \|F\|_1. \tag{27}$$

[15] See von Neuman n [11], p. 463, theorem 17.

In fact, if F' approximates to F in norm then the function $\int g(xy^{-1})\, d_y F'$ approximates uniformly to (26), and therefore it is sufficient to prove (27) for functions F which are indefinite integrals of functions $f \in C$. But our statement (27) is trivial for $h(x) = \int g(xy^{-1}) f(y)\, d_y v$.

3) If $F(E) \in AC$, then corresponding to any $\epsilon > 0$ there exists a function $g \in C$ with the properties: (i) $g(x) \geqq 0$, (ii) $M_x g(x) = 1$, and (iii) denoting the indefinite integral of (26) by $H(E)$, then

$$\| F - H \|_1 \leqq \epsilon. \tag{28}$$

In fact let F' be the indefinite integral of a function $f' \in C$ such that $\| F' - F \|_1 \leqq \epsilon/3$, and let, by 1), $g(x)$ be such that $g(x) \geqq 0$, $M_x g(x) = 1$, and

$$\int | f'(x) - M_y g(xy^{-1}) f'(y) |\, d_x v < \epsilon/3.$$

This leads to relation (28) if we apply 2) to the function $F - F'$ and the given $g(x)$.

Now if $\int \varphi(x)\, dF = 0$ for all $\varphi \in C$, then (26) vanishes identically, and therefore, by (28), $\| F \|_1 \leqq \epsilon$. Hence $F = 0$.

14. Orthogonal systems

Let C be an arbitrary module and $\{\varphi_\alpha(x)\}$ an arbitrary (not necessarily countable) orthonormal system of elements from C.[16] Every $f \in C$ has a Fourier expansion

$$f(x) \sim \sum_\alpha c_\alpha \varphi_\alpha(x), \qquad c_\alpha = \int \bar{\varphi}_\alpha(x) f(x)\, dv,$$

and so has every element $F \in V_1$, namely

$$F(E) \sim \sum_\alpha c_\alpha \varphi_\alpha(x), \qquad c_\alpha = \int \bar{\varphi}_\alpha(x)\, dF.$$

The Fourier series of $F(E)$ should be denoted more adequately by

$$\sum_\alpha c_\alpha \Phi_\alpha(E)$$

the set function $\Phi_\alpha(E)$ being the indefinite integral of $\varphi_\alpha(x)$.

On the one hand, if a sequence of functions $F(E)$ are convergent in norm, then the Fourier coefficients c_α are convergent (individually) for each α. On the other hand, if $f(x) \in R_2$, then by Bessel's inequality, only a countable number

[16] The properties of modules of almost periodic functions to which we will refer are assembled in van Kampen [17].

of Fourier coefficients is $\neq 0$, and their values tend to zero in any arrangement. Thus, by Theorem 9, we have

THEOREM 12. *If $F(E) \in AC$, then only a countable number of Fourier coefficients are $\neq 0$, (and they tend to 0 in every order).*

COUNTER EXAMPLE. The theorem is no longer true, if $F(E) \in V_1$, even if C is group invariant. Before constructing a counter example we observe that if $\{\varphi_\alpha(x)\}$ is an orthogonal system which is closed in C, then a formal series

$$\sum c_\alpha \varphi_\alpha(x)$$

is the Fourier series of an element $G(E)$ from V_1, with norm $\leqq M$, if and only if the relation

$$|\sum \bar{c}_\alpha a_\alpha| \leqq M \sup_x |f(x)| \tag{29}$$

holds for any *finite* sum

$$f(x) = \sum_\alpha a_\alpha \varphi_\alpha(x).$$

In fact, if $c_\alpha = \int \overline{\varphi_\alpha(x)}\, dG(E)$, then

$$|\sum \bar{c}_\alpha a_\alpha| = \left| \int f(x)\, d\bar{G} \right| \leqq \| G \|_1 \cdot \sup |f(x)|.$$

Conversely if (29) holds, then the linear functional

$$X(f) = \sum \bar{c}_\alpha a_\alpha$$

can be extended onto the whole space C and put in the form $X(f) = \int f(x)\, d\bar{G}$. Hence,

$$\bar{c}_\alpha = X(\varphi_\alpha) = \int \varphi_\alpha(x)\, d\bar{G}.$$

Now let C be the module of Bohr functions and $\varphi_\alpha(x) = e^{i\alpha x}$. If $K(x)$ is a function of bounded variation in $-\infty < x < \infty$, then the expression

$$X(f) = \int_{-\infty}^{\infty} f(x)\, dK(x) \qquad f \in C$$

obviously defines an element of V_C, and for its representatives in V_1 the Fourier coefficients are given by

$$c_\alpha = \int_{-\infty}^{\infty} e^{-i\alpha x}\, dK(x).$$

If so defined, c_α is continuous in α and therefore it is $\neq 0$ for a non-ennumerable number of values α.

15. The Riesz-Fischer theorem

Admitting set functions $F \epsilon V_2$, it is trivial that $\sum |a_n|^2 < \infty$ implies the existence of a function F such that

$$F \sim \sum a_n \varphi_n(x).$$

If we apply this to fields which are generated by group invariant modules C we obtain a partial generalization of the construction of Besicovitch to almost periodic functions on general groups.[17]

16. A uniqueness theorem

The theorem which follows may be compared with v. Neumann's theorems on the uniqueness of Haar measure.[18]

THEOREM 13. *Let C be group invariant. If an element $F(E)$ of V_C is right invariant that is if $F(Ea) = F(E)$ for every $a \epsilon \mathfrak{G}$ (or if it is left-invariant), then it is a constant multiple of the identity.*

The proof is quite simple. By assumption, for every element $g(x) \epsilon C$,

$$\int g(x)\, dF(E) = \int g(x)\, dF(Ea) = \int g(xa^{-1})\, dF(E).$$

By definition, $M_x g(x)$ is the uniform limit of the function

$$\frac{1}{n}\sum_{\nu=1}^{n} g(xa_\nu^{-1})$$

for appropriate choices of the elements a_ν. Hence we conclude

$$\int g(x)\, dF(E) = \int (M_y g(y))\, d_x F,$$

or, putting $\lambda = \int dF = F(\mathfrak{G})$,

$$\int g(x)\, dF(E) = \lambda M_x g(x) = \lambda \int g(x)\, dv(E).$$

Therefore $F(E) \simeq \lambda \cdot vE = \lambda\, |E|$, and this was our assertion.

Our theorem can be generalized considerably. We consider the function $T(a) = F(Ea^{-1})$ which is defined for $a \epsilon G$ and whose value is the element of V_C as represented by $F(Ea^{-1})$. Theorem 13 assumes that $T(a)$ is constant. We shall now consider the more general assumption that $T(a)$ is almost periodic. Since the space V_C is metric, many properties of numerical almost periodic functions carry over to the function $T(a)$.[19] The major property which we will consider is that it has a Fourier series by which it is uniquely determined. If

(30) $$\{\varphi_{\rho\sigma}(x)\}_{\rho,\sigma=1,\cdots,s}$$

[17] A detailed discussion will be given in Section V.

[18] See von Neumann [10] and [12].

[19] See Bochner-von Neumann [4].

denotes an irreducible representation of $\mathfrak{G}$ of dimension s consisting of almost periodic functions, then there exists a system of elements $F_{\rho\sigma}(E)$ all belonging to V_1 and hence to V_C, such that the sum

$$s \sum_{\rho,\sigma=1}^{s} \varphi_{\rho\sigma}(a) F_{\rho\sigma}(E) \tag{31}$$

is the contribution of the matrix (30) towards the Fourier series of $T(a)$. We now have the following theorem.

THEOREM 14. *Let C be group invariant. If an element $F(E) \in V_C$ is such that the function*

$$T(a) = F(Ea^{-1}) \tag{32}$$

is almost periodic, if the functions $\varphi_{\rho\sigma}(x)$ of the irreducible representation (30) *belong to C, and if* (31) *is the contribution of* (30) *to the Fourier series of* (32), *then there exist numbers $\lambda_{\rho\sigma}$ such that*

$$F_{\rho\sigma}(E) \simeq \sum_{\tau=1}^{s} \lambda_{\rho\tau} \int_E \overline{\varphi_{\tau\sigma}(x)}\, dv \tag{33}$$

for $\rho, \sigma = 1, \cdots, s$. If the functions $\{\varphi_{\rho\sigma}(x)\}$ do not belong to C then the corresponding elements $F_{\rho\sigma}(E)$ are equivalent to 0.

Obviously, if $F(Ea^{-1})$ is constant, and $\varphi(x) = 1$ is the identical representation of the group, then the term of $T(a)$ corresponding to that representation is $\varphi(a) \cdot F(E) = F(E)$, and hence we conclude from theorem 13 that $F(E) \simeq \lambda \int_E 1 \cdot dv = \lambda \mid E \mid$, which is the assertion of theorem 13 for this case.

For the proof of (33) we use the fact that $F_{\rho\sigma}(E)$ is the limit, in the norm of V_C, of sums

$$\frac{1}{n} \sum_{\nu=1}^{n} F(Ea_\nu^{-1}) \overline{\varphi_{\rho\sigma}(a_\nu)}$$

for appropriate elements a_ν. Hence, for a fixed element $g(x) \in C$, $\int g(x)\, dF_{\rho\sigma}(E)$ is the limit of

$$\int \left(\frac{1}{n} \sum_{\nu=1}^{n} g(xa_\nu) \overline{\varphi_{\rho\sigma}(a_\nu)} \right) dF(E).$$

The constants a_ν can be chosen in such a way that the integrand approximates uniformly in x, to

$$g_{\rho\sigma}(x) = M_a(g(xa) \overline{\varphi_{\rho\sigma}(a)}).$$

Hence we obtain

$$\int g(x)\, dF_{\rho\sigma}(E) = \int g_{\rho\sigma}(x)\, dF(E).$$

Putting $xa = b$, we obtain

$$g_{\rho\sigma}(x) = M_b(g(b)\bar{\varphi}_{\rho\sigma}(x^{-1}b)) = \sum_{\tau=1}^{s} \bar{\varphi}_{\rho\tau}(x^{-1})M_b(g(b)\bar{\varphi}_{\tau\sigma}(b)).$$

Hence, defining

$$\lambda_{\rho\tau} = \int \bar{\varphi}_{\rho\tau}(x^{-1})\, dF(E),$$

we will have

$$\int g(x)\, dF_{\rho\sigma}(E) = \sum_{\tau=1}^{s} \lambda_{\rho\tau} \int g(x)\bar{\varphi}_{\tau\sigma}(x)\, dv,$$

for all $g(x) \in C$; but this implies (33).

If $\varphi_{\tau\sigma}(x)$ does not belong to C, then

$$\int g(x)\bar{\varphi}_{\tau\sigma}(x)\, dv$$

vanishes, and therefore, $F_{\rho\sigma}(E) \simeq 0$.

17. A criterion for absolute continuity

From the last theorem we can easily conclude

THEOREM 15. *Let C be group invariant. If an element $F(E) \in V_C$ is such that the function* (32) *is almost periodic, then $F(E) \in V_{AC}$.*

Since (32) is almost periodic it can be approximated (uniformly in a) by finite linear combinations of expressions (31) with numerical coefficients. Putting $a = 1$, we see that $F(E)$ is the limit, in the norm of V_C, of finite linear combinations of elements $F_{\rho\sigma}(E)$ with constant coefficients. But, by the last theorem, $F_{\rho\sigma}(E)$ is the indefinite integral of an element from C.

IV. GENERALIZATION OF A THEOREM OF M. RIESZ

18. Symmetric orthonormal systems

Let $f(x)$ be a real integrable function of one variable in $0 \leqq x < 2\pi$, and let its Fourier series be denoted by

$$a_0 + \sum_{\nu=1}^{\infty} (a_\nu e^{i\nu x} + a_{-\nu} e^{-i\nu x}).$$

Let $\tilde{f}(x)$ be the function whose Fourier series is

$$-i \sum_{\nu=1}^{\infty} (a_\nu e^{i\nu x} - a_{-\nu} e^{-i\nu x})$$

provided the function exists, and let $f^*(x)$ be the function $f + i\tilde{f}$ whose Fourier series is

$$a_0 + 2 \sum_{\nu=1}^{\infty} a_\nu e^{i\nu x}.$$

The theorem of Riesz we are interested in states that if $f(x)$ belongs to the Lebesgue class L_p, $p > 1$, then the function $\tilde{f}(x)$ (and hence $f^*(x)$) also exists and is again a function of L_p.

The known proofs of the theorem are based on properties of analytic functions of a complex variable, the function

$$a_0 + 2\sum_{\nu=1}^{\infty} a_\nu e^{\nu(y+ix)}$$

being analytic in the half plane $y < 0$.[20] In the present section we shall outline a proof which is not so restricted, and which applies to orthonormal systems other than $\{e^{i\nu x}\}$.

We assume that in an arbitrary module C there exists a system $\{\varphi_\alpha(x)\}$ which is orthonormal on the generated field $\mathfrak{F}$ and has the following properties.

1) One of the given functions is the constant 1, and it will be denoted by $\varphi_0(x)$. Thus $\varphi_0(x) = 1$.

2) The other functions fall into two families. If a function $\varphi_\alpha(x)$ belongs to the first family we shall write symbolically $\alpha > 0$; if it belongs to the second family we shall put $\alpha < 0$. Corresponding to any element $\varphi_\nu(x)$ of each of the two families there exists precisely one element of the other family, which will also be denoted $\varphi_{-\nu}(x)$, such that

$$\varphi_{-\nu}(x) = \overline{\varphi_\nu(x)}.$$

3) Finally, and this is the crucial property, if $\varphi_\lambda(x)$ and $\varphi_\mu(x)$ belong to the same family, then their product $\varphi_\lambda(x)\cdot\varphi_\mu(x)$ is a *finite* linear combination

$$\sum c_\nu \varphi_\nu(x),$$

with constant coefficients c_ν, of elements $\varphi_\nu(x)$ belonging to the same family; or, more generally, it is the uniform limit of such combinations.

An orthonormal system with these properties will be called a *symmetric* system.

Obviously the system $\{e^{inx}\}$ is symmetric, an element belonging to one family if n is positive, and to the other family if n is negative, since $e^{imx}\cdot e^{inx} = e^{i(m+n)x}$. The exponent n may be an integer as in the case of periodic functions, or a dense or continuous parameter as in the case of almost periodic functions.

We can also consider the case of several variables, say $e^{i(mx+ny)}$. We obtain two families if we take in the (m, n)-plane any fixed line passing through the origin, say $\rho m + \sigma n = 0$, and if we put into one family those pairs (m, n) for which $\rho m + \sigma n > 0$ and into the other family those for which $\rho m + \sigma n < 0$. The pairs which lie on the line itself, form themselves a one-dimensional orthogonal system, and fall into two sub-families as described above, excepting always the pair $(0, 0)$. Adding each sub-family to one of the larger families, will make the total orthonormal system symmetric. The pairs (m, n) may

---

[20] Compare Zygmund [18], p. 192.

again be any two-dimensional lattice which is representative of a module of (almost) periodic functions in two variables.

The extension of the construction to more than two variables is immediate.

19

We shall now formulate two theorems.

THEOREM 16. *If $\{\varphi_\alpha(x)\}$ is a symmetric orthonormal system, and if $\{a_\alpha\}$ is a system of numbers of which only a finite number are $\neq 0$, then for $p \geqq 2$*

$$\Big\|\sum_{\alpha>0} a_\alpha \varphi_\alpha\Big\|_p \leqq A_p \Big\|\sum_{\alpha} a_\alpha \varphi_\alpha\Big\|_p \tag{34}$$

the constant A_p depending exclusively on p.

If for $q > 2$, the system $\{\varphi_\alpha\}$ is closed in R_q, then (34) also holds for $p = q/q - 1$.

Denoting by A_p an absolute constant which is not always the same we easily see that relation (34) implies and is implied by a similar relation in which $\alpha > 0$ is replaced by $\alpha \leqq 0$, or $\alpha < 0$, or $\alpha \leqq 0$. Putting

$$f(x) = a_0\varphi_0 + \sum_{\alpha>0} (a_\alpha\varphi_\alpha + a_{-\alpha}\varphi_{-\alpha}) \tag{35}$$

then

$$\begin{aligned} f + \bar{f} &= a_0 + \bar{a}_0 + \sum_{\alpha>0} ((a_\alpha + \overline{a_{-\alpha}})\varphi_\alpha + (a_{-\alpha} + \bar{a}_\alpha)\varphi_{-\alpha}) \\ f - \bar{f} &= a_0 - \bar{a}_0 + \sum_{\alpha>0} ((a_\alpha - \overline{a_{-\alpha}})\varphi_\alpha + (a_{-\alpha} - \bar{a}_\alpha)\varphi_{-\alpha}). \end{aligned} \tag{36}$$

Thus, by (34),

$$\Big\|\sum_{\alpha>0} (a_\alpha + \overline{a_{-\alpha}})\varphi_\alpha\Big\| \leqq A_p \|f + \bar{f}\| \leqq 2A_p \|f\|_p \tag{37}$$

$$\Big\|\sum_{\alpha>0} (a_\alpha - \overline{a_{-\alpha}})\varphi_{-\alpha}\Big\| \leqq A_p \|f - \bar{f}\| \leqq 2A_p \|f\|_p. \tag{38}$$

Conversely, (37) and (38) implies (34). Therefore it is sufficient to prove our theorem for real functions (35). In this case we can write

$$f = a_0 + \sum_{\alpha>0} (a_\alpha\varphi_\alpha + \bar{a}_\alpha\varphi_{-\alpha}) \qquad (a_0 \text{ real}).$$

Introducing the functions

$$\tilde{f} = -i \sum_{\alpha>0} (a_\alpha\varphi_\alpha - \bar{a}_\alpha\varphi_{-\alpha})$$

$$f^* = f + i\tilde{f} = a_0 + 2\sum_{\alpha>0} a_\alpha\varphi_\alpha$$

$$f_* = \sum_{\alpha>0} a_\alpha\varphi_\alpha$$

we have

$$2f_* = f^* - a_0,$$

and since (34) is equivalent with $\|f_*\|_p \leqq A_p \|f\|_p$ and $|a_0| \leqq \int |f(x)\varphi_0(x)| \, dv \leqq \|f\|_p$, it is sufficient to show that $\|f^*\|_p \leqq A_p \|f\|_p$, and this again is equivalent with

$$(39) \qquad \|\tilde{f}\|_p \leqq A_p \|f\|_p .$$

Now if p is a positive integer, then f_*^p is a uniform limit of linear combination

$$\sum_{\alpha>0} c_\alpha \varphi_\alpha .$$

Since $\varphi_\alpha(x)$ for $\alpha = 0$ is orthogonal to $\varphi_0(x) = 1$, we hence conclude that $M_x(f_*^p) = 0$, and consequently that

$$(40) \qquad M_x(f^{*p}) = a_0^p .$$

We now assume that p is an even integer $2k$ $(k = 1, 2, \cdots)$. The real part of (40) is

$$M_x(\tilde{f}^{2k}) - \binom{2k}{2} M_x(\tilde{f}^{2k-2} f^2) + \binom{2k}{4} M_x(\tilde{f}^{2k-4} f^4) - \cdots = (-1)^k a_0^{2k}.$$

Since, for $0 < r < s$, the Holder inequality

$$|M_x(\tilde{f}^r f^{s-r})| \leqq (\|\tilde{f}\|_s)^r \cdot (\|f\|_s)^{s-r}$$

holds, and $|a_0^{2k}| \leqq (\|f\|_{2k})^{2k}$, we readily conclude that the quotient

$$Y = \|\tilde{f}\|_{2k} : \|f\|_{2k}$$

does not exceed the largest root of the equation

$$Y^{2k} - \binom{2k}{2} Y^{2k-2} - \binom{2k}{4} Y^{2k-4} - \cdots - 2 = 0,$$

and this proves (34) for $p = 2k$.

Putting

$$K(x, y) = -i \sum_{\alpha>0} \gamma_\alpha(\varphi_\alpha(x)\overline{\varphi_\alpha(y)} - \varphi_{-\alpha}(x)\overline{\varphi_{-\alpha}(y)})$$

where $\gamma_\alpha = 1$ if $a_\alpha \neq 0$, and $\gamma_\alpha = 0$ if $a_\alpha = 0$ we have

$$\tilde{f}(x) = M_y(K(x, y)f(y)).$$

Relation (39) can now be easily completed for $2k \leqq p \leqq 2(k+1)$ by a fundamental inequality which is also due to M. Riesz.[21]

In case $1 < p < 2$ we consider any real finite sum

$$g(x) = b_0\varphi_0 + \sum_{\alpha>0} (b_\alpha\varphi_\alpha + \overline{b_\alpha}\varphi_{-\alpha})$$

[21] See Zygmund [18], p. 192.

and the corresponding function $\bar{g}(x)$. Obviously

$$| M_x(g\bar{f}) | = | M_x(\bar{g}f) | \leqq || \bar{g} ||_q \cdot || f ||_p \leqq A_q || g ||_q \cdot || f ||_p .$$

Since, by assumption, $\{\varphi_\alpha\}$ is closed in R_q, the relation

$$| M_x(g\bar{f}) | \leqq A_q || g ||_q \cdot || f ||_p$$

will hold for any $g \in R_q$. The statement of our theorem follows now if we put $g(x) = \text{sign } \bar{f}(x) \cdot | \bar{f}(x) |^{p-1}$.

By a similar argument we now obtain

THEOREM 17. *If $\{\varphi_\alpha(x)\}$ is symmetric and closed in V_p, $p > 1$, and for $1 < p < 2$ also closed in V_q, $q = p/p - 1$, then corresponding to any function*

$$F(E) \sim \sum_\alpha a_\alpha \varphi_\alpha$$

in V_p there exists a function

$$F^*(E) \sim \sum_{\alpha > 0} a_\alpha \varphi_\alpha$$

which again belongs to V_p, and $|| F^* ||_p \leqq A_p || F ||_p$.

20. Special cases

If C is the module of Bohr functions and $\varphi_\alpha(x) = e^{i\alpha x}$, then the Fourier series

$$\sum_{\alpha > 0} a_\alpha e^{i\alpha x} \tag{41}$$

of any function $F(E) \in V_p$, $p > 1$ can be "bisected" within the space V_p into the parts $\sum_{\alpha > 0}$ and $\sum_{\alpha \leqq 0}$. If β is any fixed real number, then $\sum_\alpha a_\alpha e^{i(\alpha - \beta)x}$ is again a Fourier series, the corresponding function being

$$G(E) = \int_E e^{-i\beta x} \, dF(E).$$

Hence the Fourier series (41) can be bisected at any point β, thus giving rise to the Fourier series $\sum_{\alpha > \beta}$ and $\sum_{\alpha \leqq \beta}$. Also, for any numbers β and γ,

$$\Big\| \sum_{\beta \leqq \alpha \leqq \gamma} a_\alpha e^{i\alpha x} \Big\|_p \leqq A_p \Big\| \sum_\alpha a_\alpha e^{i\alpha x} \Big\|_p .$$

In the case of almost periodic Fourier series of two variables

$$\sum_{\alpha, \beta} a_{\alpha\beta} e^{i(\alpha x + \beta y)} \tag{42}$$

belonging to V_p, we conclude that they can be cut along any straight line $\rho\alpha + \sigma\beta = 0$ and, by translation of the origin, along any line $\rho\alpha + \sigma\beta = \tau$ with arbitrary ρ, σ, τ. If in the original series (42) only integers α, β occur then the two parts of the series will also contain only integer exponents and hence belong to the Lebesgue class L_p on the torus $0 \leqq x < 2\pi$, $0 \leqq y < 2\pi$. Hence we conclude that the Fourier series

$$\sum_{m,n} a_{m,n} e^{i(mx + ny)}$$

of functions $f(x, y)$ belonging to L_p, $p > 1$, can be bisected along any line $\rho m + \sigma n = \tau$ in the (m, n)-plane and not only along the lines passing through the origin.

V. Functions of Bounded Variation in $-\infty < x < \infty$

21. Besicovitch functions

Throughout the present section C will denote the module of Bohr functions in $-\infty < x < \infty$, and all set functions will be defined on the field $\mathfrak{F}$ which is generated by this module. Therefore, if $f \in C$, its mean value $M_x f(x)$ is given by

$$\lim \frac{1}{2T} \int_{-T}^{T} f(t)\, dt \qquad (T \to \infty)$$

and if

$$\sum_{\alpha} a_\alpha e^{i\alpha x} \tag{43}$$

is the Fourier series of $f(x)$ then

$$a_\alpha = \lim \frac{1}{2T} \int_{-T}^{T} e^{-i\alpha t} f(t)\, dt, \qquad -\infty < \alpha < \infty. \tag{44}$$

According to our general definition we have $\| f \|_p = (M_x \mid f(x) \mid^p)^{1/p}$. Any e quence of elements $\{f_n\}$ from C for which

s

$$\lim \| f_m - f_n \| = 0 \qquad (m, n \to \infty), \tag{45}$$

determines a limit element belonging to V_p $(p > 1)$ or AC respectively, and the limit element has again a Fourier series of the form (43). According to our general theory the limit element is a set function $F(E)$. However in the special case we are dealing with the theory of Besicovitch enables us to "represent" our element $F(E)$ by a point function $f(x)$. In fact if the elements $f_n(x)$ from C satisfy relation (45) and if (43) denotes the term-by-term limit of the Fourier series of $f_n(x)$, then there exists a function $f(x)$ which is Lebesgue integrable over any finite interval and for which (44) holds for all real α. Also

$$\| F(E) \|_p = \left(\lim \frac{1}{2T} \int_{-T}^{T} \mid f(t) \mid^p dt \right)^{1/p}.$$

We now start, conversely, from an "arbitrary" function $f(x)$ which is integrable over any finite interval and for which the limit (44) exists for every real α; or more generally, from an arbitrary function $F(x)$ which is of bounded variation over any finite interval and for which the limit

$$a_\alpha = \lim \frac{1}{2T} \int_{-T}^{T} e^{-i\alpha t}\, dF(t) \tag{46}$$

exists for all real α; and we set up the series (43) with these coefficients. The series (43) is not a Fourier series by definition. However, under fairly general

conditions there exists a set function $F(E)$ whose Fourier series coincides with (43).

THEOREM 18. *In order that (43) is the Fourier series of a function $F(E)$ belonging to V_1 it is necessary and sufficient that for every exponential polynomial*

$$g(x) = \sum_{\alpha} b_\alpha e^{i\alpha x} \tag{47}$$

the relation

$$\left| \lim \frac{1}{2T} \int_{-T}^{T} \overline{g(t)}\, dF(t) \right| \leqq M \cdot \sup_x |g(x)| \tag{48}$$

holds. In particular it is sufficient that

$$\overline{\lim} \frac{1}{2T} \int_{-T}^{T} |dF(t)| \leqq M. \tag{49}$$

In order that $F(E)$ belongs to AC it is sufficient that, moreover, the function

$$e(h) = \sup_{|g| \leqq 1} \left| \lim \frac{1}{2T} \int_{-T}^{T} \overline{g(t)}\, d_t[F(t+h) - F(t)] \right| \tag{50}$$

is almost periodic in $-\infty < h < \infty$. In particular it is sufficient that, moreover, the function

$$e_0(h) = \overline{\lim} \frac{1}{2T} \int_{-T}^{T} |d_t[F(t+h) - F(T)]| \tag{51}$$

is continuous at $h = 0$ that is $\lim_{h \to 0} e_0(h) = 0$.

In order that $F(E)$ belongs to V_p it is necessary and sufficient that

$$\left| \lim \frac{1}{2T} \int_{-T}^{T} \overline{g(t)}\, dF(t) \right| \leqq M \cdot \| g \|_q, \qquad q = p/p - 1. \tag{52}$$

In particular, if $F(x) = \int^x f(x)\, dx$ it is sufficient that

$$\overline{\lim} \frac{1}{2T} \int_{-T}^{T} |f(t)|^p\, dt \leqq M^p.$$

PROOF. We remarked in §14 that (43) is the Fourier series of a function $F(E) \in V_1$ if and only if for any exponential polynomial (47) the relation

$$\left| \sum_{\alpha} \bar{b}_\alpha a_\alpha \right| \leqq M \cdot \sup_x |g(x)| \tag{53}$$

holds. But, on account of (46),

$$\sum_{\alpha} \bar{b}_\alpha a_\alpha = \lim \frac{1}{2T} \int_{-T}^{T} \overline{g(t)}\, dF(t),$$

and therefore (48) is equivalent with (53).

The numerical function $e(h)$ is the "translation function" of the abstract

function $F(E + h)$ whose values are elements of V_C. The latter function is almost periodic if and only if $e(h)$ is almost periodic. Therefore, by theorem 15, $F(E)$ belongs to AC if $e(h)$ is almost periodic. $e(h)$ obviously has the properties

$$e(h) \geqq 0, \qquad e(0) = 0, \qquad e(h_1 + h_2) \leqq e(h_1) + e(h_2). \tag{54}$$

Furthermore, $e(h) \leqq e_0(h)$. Therefore, if $e_0(h)$ is continuous at the origin, then so is $e(h)$. But a function $e(h)$ with the properties (54) which is continuous at the origin is almost periodic,[22] and this proves our statement involving the function $e_0(h)$.

Finally the last part of the theorem is a consequence of the fact that a function $F(E)$ belongs to V_p if it represents a functional on V_q.

22. A counter example

By the theory of Besicovitch all functions $F(E)$ which belong to AC can be represented by point functions. We are going to show that this is no longer true for all functions which belong to V_1. If the Fourier coefficient a_α is defined by (46) or by some similar formula it is Lebesgue measurable as a function of α. However we shall exhibit a function $F(E)$ whose Fourier coefficient is not so measurable.

Let $\{\xi_r\}$ be a Hamel basis for real numbers. Any real number α can be represented uniquely in the form $\sum_r p_r\xi_r$ with rational coefficients p_r of which only a finite number are $\neq 0$. If (47) is an exponential polynomial we may (and we shall) assume that the index α ranges over a lattice

$$\alpha = \sum_{r=1}^{k} n_r \frac{\xi_r}{N}$$

where N is a suitable positive integer and the symbols $n_1, \cdots, n_k$ range over all real integers. We now take an arbitrary function $\eta_r = \lambda(\xi_r) > 0$ and we put

$$A(\alpha) = \sum_{r=1}^{k} |n_r| \frac{|\eta_r|}{N}$$

and

$$a_\alpha = e^{-A(\alpha)}. \tag{55}$$

For $t > 0$ the series

$$\sum_{n=-\infty}^{\infty} \exp\{-|n|\,t + inx\}$$

converges absolutely and uniformly in $-\infty < x < \infty$ towards a positive sum, and therefore the expression

$$\begin{aligned} P(x) &= \sum_{\alpha} e^{-A(\alpha)} e^{i\alpha x} \\ &= \prod_{r=1}^{k} \sum_{n_r=-\infty}^{\infty} \exp\left\{-|n_r| \frac{|\eta_r|}{N} + in_r \frac{\xi_r x}{N}\right\} \end{aligned}$$

defines a positive almost periodic function. It is easy to verify that $M_x P(x) = 1$.

---

[22] See Bochner [3], p. 136–7.

Now

$$\sum_\alpha \bar{b}_\alpha a_\alpha = M_x \left(\sum_\alpha e^{-A(\alpha)} e^{i\alpha x}\right) \overline{g(x)})$$

and therefore

$$\left| \sum_\alpha \bar{b}_\alpha a_\alpha \right| \leqq M_x P(x) \cdot \sup_x | g(x) | = \sup | g(x) |,$$

and hence we conclude that the expression (55) is the Fourier coefficient of a function $F(E)$ belonging to V_1. However it is possible to choose the function $\eta_r = \lambda(\xi_r)$ in such a way that the resulting function $A(\alpha)$ is not Lebesgue measurable in α.[23] Q.E.D.

23. Linearly independent exponents

Finally we shall briefly consider the case in which the exponents α for which the coefficient a_α is different from zero are linearly independent; in which case the series (43) can be written in the form

$$\sum_r a(\xi_r)\, e^{i\xi_r x} \tag{56}$$

the system $\{\xi_r\}$ being an appropriate Hamel basis. Putting $\varphi(\xi, x) = e^{i\xi x}$ and assuming that $\eta_1 \leqq \eta_2 \cdots \leqq \eta_m$, $\zeta_1 \leqq \zeta_2 \cdots \leqq \zeta_n$, then the integral

$$M_x \varphi(\eta_1, x) \cdots \varphi(\eta_m, x) \cdot \overline{\varphi(\zeta_1, x)} \cdots \overline{\varphi(\zeta_n, x)} \tag{57}$$

has the value 1 if $m = n$ and $\eta_r = \zeta_r$, $r = 1, \cdots, m$, and the value 0 in all other cases. Consequently, any *finite* sum

$$f(x) = \sum_{r=1}^{s} a(\xi_r) \varphi(\xi_r, x)$$

satisfies for $k = 1, 2, 3, \cdots$ the inequality

$$\| f \|_{2k} \leqq (k!)^{1/2k} \| f \|_2 . \tag{58}$$

In fact,[24] if we substitute the values of (57) we obtain

$$M_x | f(x) |^{2k} = \sum \left(\frac{(n_1 + \cdots + n_s)!}{n_1! \cdots n_s!} | a(\xi_1) |^{n_1} \cdots | a(\xi_s) |^{n_s} \right)^2$$

the sum ranging over $n_1 \geqq 0, \cdots, n_s \geqq 0$, $n_1 + \cdots + n_s = k$. Therefore

$$M_x | f(x) |^{2k} \leqq k! \left(\sum_{r=1}^{s} | a(\xi_r) |^2 \right)^k = k! \| f \|_2^{2k} .$$

Combining (58) with Holder's inequality we obtain

$$\| f \|_2 \leqq \| f \|_1^{2/3} \cdot \| f \|_4^{1/3} \leqq 2 \| f \|_1^{2/3} \cdot \| f \|_2^{1/3}$$

[23] See Sierpinski [11].

[24] Compare Kaczmarz-Steinhaus [8], p. 131.

and thus

$$\|f\|_2 \leqq 4\,\|f\|_1. \tag{59}$$

If (56) is the Fourier series of $F(E)$ we choose a finite number of elements $\xi_1, \cdots, \xi_s$ of the Hamel basis and we form the kernel

$$K_n(x) = \prod_{r=1}^{s} \sum_{\nu=-n}^{n} \left(1 - \frac{|\nu|}{n}\right) e^{i\nu\xi_r x}$$

and with this kernel the point function

$$f(x) = \int_{\mathfrak{G}} K_n(x-t)\, d_t F(E).$$

Since $K_n(x) \geqq 0$ and $M_x K_n(x) = 1$, it is easily seen that

$$\|f\|_1 \leqq \|F\|_1.$$

On the other hand the Fourier series of $f(x)$ is

$$\frac{n-1}{n} \sum_{r=1}^{s} a(\xi_r)\, e^{i\xi_r x}$$

and therefore by (59),

$$\left(\frac{n-1}{n}\right)^2 \sum_{r=1}^{s} |a(\xi_r)|^2 \leqq 16\,\|F\|_1^2.$$

The integer n being arbitrary we finally obtain

$$\sum_r |a(\xi_r)|^2 \leqq 16\,\|F\|_1^2.$$

In particular we may draw the following conclusion.

THEOREM 19. *If $F(x)$ is of bounded variation on every finite interval, if* (49) *holds, if the limit* (46) *holds for all real α, and if the indices α for which $a_\alpha \neq 0$ are linearly independent, then the set of these indices is countable, and the series*

$$\sum_\alpha |a_\alpha|^2$$

is convergent.

PRINCETON UNIVERSITY.

BIBLIOGRAPHY

[1] S. Banach, *Opérations linéaires*, 1932.
[2] A. S. Besicovitch, *Almost periodic functions*, Cambridge, 1932.
[3] S. Bochner, "Beiträge zur Theorie der fastperiodischen Funktionen," *Math. Annalen*, 96 (1926), 119–147.
[4] S. Bochner and J. v. Neumann, "Almost periodic functions in groups," *Trans. Amer. Math. Soc.*, 37 (1935), 21–50.

[5] H. Bohr, "Kleinere Beiträge zur Theorie der fastperiodischen Funktionen II," *Danske Videnskabernes Selskab*, X, 10 (1931).
[6] P. J. Daniell, "A general form of integral," *Annals of Math.*, 19 (1917/18), 279–294.
[7] Borge Jessen, "Abstract Maal-og Integralteori I," *Matematisk Tidskrift*, B 1934, 73–84.
[8] S. Kaczmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, 1935.
[9] A. P. Morse, "Convergence in variation and related topics," *Trans. Amer. Math. Soc.*, 41 (1937), 48–83.
[10] J. v. Neumann, "Zum Haarschen Mass in topologischen Gruppen," *Compositio Mathematica*, 1 (1934), 106–114.
[11] J. v. Neumann, "Almost periodic functions in a group," *Trans. Amer. Math. Soc.*, 36 (1934), 445–492.
[12] J. v. Neumann, "The uniqueness of Haar's measure," *Recueil Mathématique de Moscou*, 1 (1936), 721–734.
[13] A. Plessner, "Eine Kennzeichnung der totalstetigen Funktionen," *Journal für reine und angewandte Mathematik*, 160 (1929), 26–32.
[14] F. Riesz, "Su alcune disugualianze," *Bolletino dell'Unione Matematica Italiana* 2 (1928), 1–3.
[15] S. Saks, *Theory of the Integral*, 1937.
[16] W. Sierpinski, "Sur une propriété des fonctions de M. Hamel," *Fundamenta Matematicae*, 5 (1924), 334–8.
[17] E. R. Van Kampen, "Almost periodic functions and compact groups," *Annals of Math.*, 37 (1936), 78–91.
[18] A. Zygmund, *Trigonometrical series*, 1935.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

ON A NECESSARY CONDITION FOR THE STRONG LAW OF LARGE NUMBERS

By Paul R. Halmos

(Received April 3, 1939)

Let $x_1, x_2, \cdots$ be a sequence of independent chance variables. Define the chance variable $\bar{x}_n$ to be x_n whenever $|x_n| \leqq n$, and to be zero otherwise (for $n = 1, 2, \cdots$). Write $b_n = E[(\bar{x}_n - E(\bar{x}_n))^2]$.[1] We shall suppose that

$$\lim_{n\to\infty} \frac{E(\bar{x}_1) + \cdots + E(\bar{x}_n)}{n} = 0. \tag{1}$$

The following theorem concerning these chance variables is an easy consequence of a result due to Kolmogoroff.[2]

Theorem I. *The two conditions*

$$P\left\{\lim_{n\to\infty} \frac{x_n}{n} = 0\right\} = 1^{3} \tag{2}$$

and

$$\sum_{n=1}^{\infty} \frac{b_n}{n^2} < \infty \tag{3}$$

imply that

$$P\left\{\lim_{n\to\infty} \frac{x_1 + \cdots + x_n}{n} = 0\right\} = 1. \tag{4}$$

Proof: Since $\sum_1^\infty \frac{b_n}{n^2} < \infty$, the immediate application of Kolmogoroff's result tells us that

$$P\left\{\lim_{n\to\infty} \frac{\bar{x}_1 - E(\bar{x}_1) + \cdots + \bar{x}_n - E(\bar{x}_n)}{n} = 0\right\} = 1. \tag{5}$$

Conditions (1) and (5) imply that the sequence $\{\bar{x}_n\}$ satisfies (4). It remains to prove that this in turn implies that the sequence $\{x_n\}$ satisfies (4). This can be done by proving that the sequences $\{\bar{x}_n\}$ and $\{x_n\}$ are equivalent in the sense

[1] $E(x)$ is the expectation of the chance variable x.

[2] C. R. Acad. Sci., Paris. 191(1930) pp. 910–911.

[3] If p is a proposition, $\{p\}$is the event "p is true" and $P\{p\}$ is the probability of $\{p\}$.

of Khintchine.[4] In other words we are to prove that

$$\sum_{n=1}^{\infty} P\{x_n \neq \bar{x}_n\} < \infty. \tag{6}$$

But $P\{x_n \neq \bar{x}_n\} = P\{|x_n| > n\}$. Hence, if (6) is not satisfied, by the Borel-Cantelli lemma[5] the probability is one that an infinite number of the conditions $|x_n/n| > 1$ are satisfied. Since this is impossible, by (2), the theorem is proved.[6]

The converse of this theorem is not known. It is the purpose of this note to make a contribution in this direction, by proving the following theorem.

THEOREM II. *If* $P\left\{\frac{x_1 + \cdots + x_n}{n} \to 0\right\} = 1$, *then* $P\left\{\frac{x_n}{n} \to 0\right\} = 1$ *and*

$$\sum_{n=1}^{\infty} \frac{b_n}{n^{2+\epsilon}} < \infty \qquad \textit{for all } \epsilon > 0. \tag{7}$$

The first part of the theorem is an elementary property of $(C, 1)$ convergence. It is necessary to prove (7) only. It is clear, of course, that (7) does not imply (3). According to (7) the series in (3) may diverge, but it may not diverge too rapidy. In most cases it is not difficult to verify (7).

For the purposes of the proof we assume that the chance variables x_n are represented as measurable functions on a space Ω, so that probability corresponds to a probability measure, P, defined on a Borel field of subsets of Ω, and the expectation of the chance variable x corresponds to the integral $\int x\, dP$.[7]

PROOF OF THEOREM II. The measure of the set of points of Ω where $(\bar{x}_1 + \cdots \bar{x}_n)/n \to 0$ is easily derived from Kolmogoroff's triple limit condition for convergence almost everywhere.[8] We obtain

$$P\left\{\frac{\bar{x}_1 + \cdots + \bar{x}_n}{n} \to 0\right\} \leqq \lim_{\eta \to 0} \lim_{n \to \infty} \lim_{N \to \infty} P\left\{\left|\frac{\bar{x}_n}{n}\right| \leqq \eta, \cdots, \left|\frac{\bar{x}_n + \cdots + \bar{x}_N}{N}\right| \leqq \eta\right\}. \tag{8}$$

We consider, for the time being, $\eta > 0$, n and N as fixed. For $k = n, \cdots, N$ let F_k be the set where

$$\left|\frac{\bar{x}_n}{n}\right| \leqq \eta, \cdots, \left|\frac{\bar{x}_n + \cdots + \bar{x}_k}{k}\right| \leqq \eta. \tag{9}$$

---

[4] Fréchet, M. *Généralités sur les probabilités. Variables aléatoires.* Paris, 1937. p. 251.

[5] Ibid. p. 27.

[6] Actually Kolmogoroff's result can be stated slightly more generally, but it is here more convenient to state theorem I in its present form in order better to exhibit its relation to theorem II.

[7] For a discussion of the representation of sequences of chance variables see J. L. Doob, *Stochastic processes with an integral-valued parameter.* Trans. of the A. M. S. 44(1938) pp. 87–95. See also P. R. Halmos, *Invariants of certain stochastic transformations: the mathematical theory of gambling systems.* Duke Mathematical Journal. 5(1939) pp. 461–478.

[8] Math. Annalen. 99(1928) p. 315.

Then $F_n \supseteq F_{n+1} \supseteq \cdots \supseteq F_{N-1} \supseteq F_N$, and if we write $E_n = CF_n$, $E_{n+1} = F_n \cdot CF_{n+1}, \cdots, E_N = F_{N-1}CF_N$,[9] then we have

$$P(F_N) + \prod_{k=n}^{N} P(E_k) = 1. \tag{10}$$

We define, for $k = n, \cdots, N$

$$a_k = \int (z_n + \cdots + z_k) dP_{F_k} = \frac{1}{P(F_k)} \int_{F_k} \sigma_k \, dP, \tag{11}$$

where $z_k = \bar{x}_k - \int \bar{x}_k dP$ and $\sigma_k = z_n + \cdots + z_k$.[10] (We shall also write $s_k = \bar{x}_n + \cdots + \bar{x}_k$ for $k = n, \cdots, N$; $\sigma_{n-1} = s_{n-1} = 0$). Then, on the set F_k, $(k = n, \cdots, N)$,

$$\begin{aligned} | \sigma_k - a_k | &= \left| s_k - \int s_k \, dP - \int s_k \, dP_{F_k} + \int \left\{ \int s_k \, dP \right\} dP_{F_k} \right| \\ &= \left| s_k - \int s_k \, dP_{F_k} \right| \leqq | s_k | + \frac{1}{P(F_k)} \left| \int_{F_k} s_k \, dP \right| \\ &\leqq k\eta + \frac{1}{P(F_k)} \cdot k\eta \cdot P(F_k) = 2k\eta. \end{aligned} \tag{12}$$

If we set $a_{n-1} = 0$, we have for $k = n, \cdots, N$

$$\begin{aligned} | a_k - a_{k-1} | &= \left| \int \sigma_k \, dP_{F_k} - \int \sigma_{k-1} \, dP_{F_{k-1}} \right| \\ &\leqq \left| \int z_k \, dP_{F_k} \right| + \left| \int s_{k-1} dP_{F_k} - \int s_{k-1} \, dP - \int s_{k-1} \, dP_{F_{k-1}} + \int s_{k-1} \, dP \right| \\ &\leqq 2k + \left| \int s_{k-1} \, dP_{F_k} \right| + \left| \int s_{k-1} \, dP_{F_{k-1}} \right| \\ &= 2k + \frac{1}{P(F_k)} \left| \int_{F_k} s_{k-1} \, dP \right| + \frac{1}{P(F_{k-1})} \left| \int_{F_{k-1}} s_{k-1} \, dP \right| \\ &\leqq 2k + (k-1)\eta + (k-1)\eta \leqq 2k(1+\eta). \end{aligned} \tag{13}$$

Similarly, for $k = n, \cdots, N$, we have, (setting $F_{n-1} = 0$),

$$\begin{aligned} \int_{F_{k-1}} (\sigma_k - a_k)^2 \, dP &= \int_{F_k} + \int_{E_k} (\sigma_k - a_k)^2 \, dP \\ &= \int_{F_k} (\sigma_k - a_k)^2 \, dP + \int_{E_k} [(\sigma_{k-1} - a_{k-1}) - (a_k - a_{k-1}) + z_k]^2 \, dP \\ &\leqq \int_{F_k} (\sigma_k - a_k)^2 \, dP + \int_{E_k} [2(k-1)\eta + 2k(1+\eta) + 2k]^2 \, dP^{11} \\ &\leqq \int_{F_k} (\sigma_k - a_k)^2 \, dP + P(E_k)(4k(1+\eta))^2. \end{aligned} \tag{14}$$

---

[9] CE is the complement in Ω of the set E.

[10] We note that since $| \bar{x}_k | \leqq k$, $| z_k | \leqq 2k$.

[11] Since $E_k = F_{k-1} \cdot CF_k \subseteq F_{k-1}$.

Finally, for $k = n, \cdots, N$,

$$
\begin{aligned}
\int_{F_{k-1}} (\sigma_k - a_k)^2\, dP &= \int_{F_{k-1}} [(\sigma_{k-1} - a_{k-1}) - (a_k - a_{k-1}) + z_k]^2\, dP \\
&= \int_{F_{k-1}} [(\sigma_{k-1} - a_{k-1})^2 + (a_k - a_{k-1})^2 + z_k^2]\, dP \\
&\geqq \int_{F_{k-1}} (\sigma_{k-1} - a_{k-1})^2\, dP + P(F_{k-1}) \int z_k^2\, dP \text{ [12]} \\
&\geqq \int_{F_{k-1}} (\sigma_{k-1} - a_{k-1})^2\, dP + P(F_N) \int z_k^2\, dP.
\end{aligned} \tag{15}
$$

Combining the inequalities (14) and (15) we obtain

$$
\begin{aligned}
&\int_{F_{k-1}} (\sigma_{k-1} - a_{k-1})^2\, dP + P(F_N) \int z_k^2\, dP \\
&\qquad \leqq \int_{F_k} (\sigma_k - a_k)^2\, dP + P(E_k)\cdot 16k^2(1+\eta)^2.
\end{aligned} \tag{16}
$$

Now let ϵ be any positive number. Dividing (16) through by $k^{2+\epsilon}$ we obtain

$$
\begin{aligned}
&\frac{1}{k^{2+\epsilon}} \int_{F_{k-1}} (\sigma_{k-1} - a_{k-1})^2\, dP + P(F_N) \frac{1}{k^{2+\epsilon}} \int z_k^2\, dP \\
&\qquad \leqq \frac{1}{k^{2+\epsilon}} \int_{F_k} (\sigma_k - a_k)^2\, dP + P(E_k)\cdot 16(1+\eta)^2.
\end{aligned} \tag{17}
$$

Hence, summing (17) for $k = n, \cdots, N$,

$$
\begin{aligned}
&\sum_{k=n}^{N} \frac{1}{k^{2+\epsilon}} \int_{F_{k-1}} (\sigma_{k-1} - a_{k-1})^2\, dP + P(F_N) \sum_{k=n}^{N} \frac{b_k}{k^{2+\epsilon}} \\
&\qquad \leqq \sum_{k=n}^{N} \frac{1}{k^{2+\epsilon}} \int_{F_k} (\sigma_k - a_k)^2\, dP + 16(1+\eta)^2 \sum_{k=n}^{N} P(E_k). \text{ [13]}
\end{aligned} \tag{18}
$$

If we write $\alpha_k = \int_{F_k} (\sigma_k - a_k)^2\, dP$, then we have (if $\epsilon < 1$)

$$
\begin{aligned}
P(F_N) \sum_{k=n}^{N} \frac{b_k}{k^{2+\epsilon}} \leqq \sum_{k=n}^{N-1} \alpha_k \left(\frac{1}{k^{2+\epsilon}} - \frac{1}{(k+1)^{2+\epsilon}} \right) + \frac{\alpha_N}{N^{2+\epsilon}} \\
+ 16(1+\eta)^2 \sum_{k=n}^{N} P(E_k)
\end{aligned}
$$

[12] Since z_k is independent of F_{k-1}.

[13] $b_k = \int z_k^2\, dP$.

$$\begin{aligned}
&\leqq \sum_{k=n}^{N-1} \alpha_k \left(\frac{-k^{2+\epsilon} + k^{2+\epsilon} + (2+\epsilon)k^{1+\epsilon} + \cdots}{k^{4+2\epsilon}} \right) + P(F_N) \\
(19) \qquad & \qquad\qquad\qquad\qquad\qquad\qquad + 16(1+\eta)^2 \sum_{k=n}^{N} P(E_k) \\
&\leqq 3 \sum_{k=n}^{N-1} \alpha_k \frac{k^{1+\epsilon}}{k^{4+2\epsilon}} + 16(1+\eta)^2 \left[P(F_N) + \sum_{k=n}^{N} P(E_k) \right] \\
&= 3 \sum_{k=n}^{N-1} \alpha_k \frac{1}{k^{3+\epsilon}} + 16(1+\eta)^2 \leqq 3 \sum_{k=n}^{N-1} \frac{1}{k^{1+\epsilon}} + 16(1+\eta)^2 \\
&\leqq K,
\end{aligned}$$

where K is a constant independent of n, N, or η (as long as we take, as we may without any loss of generality, η to be bounded). This last inequality concludes the proof of (7). For if the series in (7) diverges for any ϵ, then by choosing N sufficiently large we can dominate $P(F)$ by arbitrarily small numbers. But this contradicts the assumption that $\lim_{\eta \to 0} \lim_{n \to \infty} \lim_{N \to \infty} P(F_N) = 1$. Hence theorem II is proved.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

ÜBER EINE VERALLGEMEINERUNG DER STETIGEN FASTPERIODISCHEN FUNKTIONEN VON H. BOHR

von B. Lewitan

(Received January 19, 1939)

In vorliegender Arbeit wird eine verallgemeinerung der Bohrschen Funktionen gegeben, in derem Gebiete der Eindentigkeitssatz bestehen bleibt.[1]

Die Arbeit zerfällt in zwei Teile. Im ersten Teile untersuche ich die verallgemeinerten fastperiodischen Funktionen und beweise den Eidentigkeitssatz. Dabei spielen eine grosse Rolle fastperiodische Mengen ganzer Zahlen, von welchen schon Besicovitch oft Gebrauch gemacht hat.

Im zweiten Teile beweise ich, dass beschränkte Lösungen linearer Differentialgleichungen mit fastperiodischen Koeffizienten unter genügend allgemeinen Bedingungen Funktionen sind, welche in unserem Sinne die verallgemeinerte Fastperiodizität besitzen.

Kapitel 1. Über den Eindeutigkeitssatz

1. Es sei E eine relativ dichte Menge ganzer Zahlen $\{\tau_i\}$ $(i = 0, \pm 1, \pm 2, \cdots;$ $\tau_{-i} = -\tau_i)$. Wir werden sagen, dass E bis auf η (>0) fastperiodisch ist, falls man, aus E eine solche relativ dichte Untermenge E' auswählen und eine solche Zahl $I_0 > 0$ angeben kann, dass für jedes Interval (a, b) welches länger als I_0 ist, die Punkte der Menge $E \cdot (a, b)$ nach der Verschiebung um jede Zahl $\tau \subset E'$ in Punkte von E übergehen, mit Ausnahme höchstens von $\eta(b - a)$ aus diesen Punkten.

Wir werden einfach sagen, dass E *fastperiodisch* ist, wenn E fastperiodisch ist bis auf jedes $\eta > 0$.

Es sei $f(x)$ fastperiodisch im Sinne von H. Bohr und sei ferner $\bar{E}_\epsilon$ die Menge aller ganzer ϵ-Verschiebungen der Funktion $f(x)$. Dann ist, wie bekannt,[2] $\bar{E}_\epsilon$ fastperiodisch für fast alle ϵ.

Satz I. *Der Durchschnitt von zwei fastperiodischen Mengen $E^{(1)}$ und $E^{(2)}$ ist auch eine fastperiodische Menge.*

Beweis. Es sei E eine fastperiodische Menge von ganzen Zahlen

$$E = \{\tau_i\} \qquad (i = 0, \pm 1, \pm 2, \cdots; \tau_{-i} = -\tau_i).$$

[1] Vorläufige Mitteilungen habe ich ohne Beweis in den C. R. der Akademie des U. S. S. R. publiziert (Bd. XVII N 6, p. 287–290, Bd. XIX, N 6–7, p. 447–450) Ausführliche aber auf andere Definitionen stützende Darstellung habe ich zuerst in den Communications de Sciences Math. de Kharkoff serie 4, t. XV_2, p. 3–35 (1938) gegeben. Neue Definitionen, welche ich hier benutze, erlauben die Darstellung wesentlich kürzer zu machen.

[2] S. Besicovitch, *Almost periodic functions*, Cambridge, 1932, p. 55–59.

Dann ist die Funktion

$$K_\delta(t) = \begin{cases} 1 & \text{für} \quad \tau_i \leqq t \leqq \tau_i + \delta \\ 0 & \text{für übrige } t \end{cases} \qquad (0 < \delta < 1),$$

fastperiodisch in Sinne von H. Weyl.[3]

Für jedes $\eta > 0$ gibt es, in der Tat, eine Zahl $L > 0$ und eine relativ dichte Untermenge $E' \subset E$ sodass für jedes c die Punkte von $E \cdot (c, c + L)$ nach der Verschiebung um jede Zahl aus E' wieder in Punkte von E übergehen mit Ausnahme von höchstens ηL unter den verschobenen Punkten. Folglich ist für jedes $\tau \subset E'$

$$\frac{1}{L} \int_c^{c+L} | K_\delta(t + \tau) - K_\delta(t) | \, dt \leqq 2\eta\delta < 2\eta,$$

was die Fastperiodizität von $K_\delta(t)$ im Weylschen Sinne beweist.

Umgekehrt, ist $K_\delta(t)$ fastperiodisch im Sinne von H. Weyl und bilden für jedes η die η-Verschiebungen von $K_\delta(t)$ eine Untermenge von E, so ist die Menge E fastperiodisch.

Es mögen jetzt $K_\delta^{(1)}(t)$ und $K_\delta^{(2)}(t)$ den beiden Mengen $E^{(1)}$ und $E^{(2)}$ entsprechen. Der Menge $E^{(1)} \cdot E^{(2)}$ entspricht dann offenbar die Funktion $K_\delta(t) = K_\delta^{(1)}(t) \cdot K_\delta^{(2)}(t)$ und diese Funktion ist im Sinne von H. Weyl fastperiodisch, weil die beiden Faktoren solche Funktionen sind. Die Verschiebungen von $K_\delta(t)$ sind die gemeinsamen Verschiebungen von $K_\delta^{(1)}(t)$ und $K_\delta^{(2)}(t)$ und bilden folglich eine Untermenge von $E^{(1)} \cdot E^{(2)}$. Daher ist $E^{(1)} \cdot E^{(2)}$ relativ dicht und fastperiodisch.

Definition. *Eine in* $-\infty < x < \infty$ *stetig erklärte Funktion* $f(x)$ *heisst N-fastperiodisch, wenn für beliebige* $\epsilon > 0$ *und* $N > 0$ *eine solche fastperiodische Menge ganzer Zahlen* $\tau_n(\epsilon, N)$ $(n = 0, \pm 1, \pm 2, \cdots; \tau_{-n} = -\tau_n)$ *angegeben werden kann, welche den Ungleichungen*

$$| f(x + \tau_n) - f(x) | < \epsilon, \quad | x | < N \quad (n = 0, \pm 1, \pm 2, \cdots)$$

genügen.

Nach einem bekannten Satze,[4] ist jede Funktion von H. Bohr N-fastperiodisch. Aber schon einfache Operationen mit Bohrschen Funktionen führen auf Funktionen, welche im Sinne von H. Bohr nicht mehr fastperiodisch sind, während sie N-fastperiodisch bleiben.

Es sei, z.B., $f(x)$ eine Funktion von H. Bohr, für welche

$$f(x) > 0, \quad \inf_x f(x) = 0;$$

(man könnte

$$f(x) = 2 + \cos \lambda_1 x + \cos \lambda_2 x$$

mit irrationalem λ_1/λ_2 nehmen).

[3] S. Besicovitch, p. 92.

[4] Besicovitch, p. 55-59.

Die Funktion

$$\rho(x) = \frac{1}{f(x)}$$

ist im Sinne von H. Bohr nicht fastperiodisch, weil sie nicht beschränkt ist. Aber, N-fastperiodisch ist $\rho(x)$ gewiss, denn bedeutet τ eine ϵ-Verschiebung von $f(x)$ und ist für $|x| < N$ $f(x) > \kappa$, so besteht die Ungleichung

$$|\rho(x+\tau) - \rho(x)| = \frac{|f(x+\tau) - f(x)|}{f(x+\tau)f(x)} < \frac{\epsilon}{\kappa(\kappa - \epsilon)} < \delta,$$

für $|x| < N$ und $\epsilon < \min\left(\frac{\kappa}{2}, \frac{\kappa^2\delta}{2}\right)$.

Und da für $f(x)$ die Menge aller ganzer ϵ-Verschiebungen fastperiodisch für fast alle ϵ war so ist $\rho(x)$ wirklich eine N-fastperiodische Funktion.

Satz II. *Summe und Produkt von zwei N-fastperiodischen Funktionen ist wieder eine N-fastperiodische Funktion.*

Beweis. In bezug auf die Summe folgt die Behauptung unmittelbar aus dem Satze I. Um den Satz in bezug auf das Produkt zu beweisen, genügt es zu zeigen, dass $f^2(x)$ N-fastperiodisch ist, wenn $f(x)$ N-fastperiodisch war.

Zum beweise bemerkte man, dass wenn für $|x| < N$

$$|f(x+\tau) - f(x)| < \epsilon$$

gilt,

$$|f(x+\tau)| < |f(x+\tau) - f(x)| + |f(x)| < \epsilon + g_N \quad (|x| < N).$$

Somit haben wir

$$\begin{aligned}|[f(x+\tau)]^2 - [f(x)]^2| &= |f(x+\tau) - f(x)|\,|f(x+\tau) + f(x)| \\ &\leqq \epsilon(2g_N + \epsilon) < \delta, \qquad \text{falls} \quad \epsilon < \frac{\delta}{2g_N + 1}, \qquad |x| < N.\end{aligned}$$

Insbesondere ist das Produkt einer N-fastperiodischen Funktion und einer Bohrschen Funktion eine N-fastperiodische Funktion.

Da das Produkt einer N-fastperiodischen Funktion mit einer Konstante wieder N-fastperiodisch ist so folgt aus dem Satze II, dass die Klasse der N-fastperiodischen Funktionen linear ist.

Insbensondere ist die Differenz zweier N-fastperiodischen Funktionen auch N-fastperiodisch.

2. Es sei $f(x)$ eine in jedem endlichen Intervalle messbare Funktion, und es sei

$$\text{(1)} \qquad \overline{\lim_{T\to\infty}} \frac{1}{2T}\int_{-T}^{T} |f(x)|^2\,dx < \infty.$$

Setzt man noch voraus für alle reelle λ die Existenz der Mittelwerte

$$a(\lambda) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)e^{-i\lambda x}\, dx,$$

so entspricht bekanntlich jeder Funktion $f(x)$ eine Fouriersche Reihe

$$f(x) \sim \sum A_n e^{i\Lambda_n x}$$

wobei

$$A_n = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)e^{-i\Lambda_n x}\, dx.$$

Die Zahlen A_n bezeichnen wir als Fouriersche Koeffizienten der Funktion $f(x)$, und Λ_n als Fouriersche Exponenten der Funktion $f(x)$.

Es sei nun $f(x)$ eine N-fastperiodische Funktion. Dann gilt folgender Fundamentalsatz:

Satz III. *Ist $f(x)$ eine N-fastperiodische Funktion, die der Ungleichung* (I) *genügt, und existiert für jeden reellen λ der Mittelwert $a(\lambda)$, so kann man für beliebige $\epsilon > 0$ und $N > 0$ ein trigonometrisches Polynom $P(x)$ angeben, das die Ungleichung*

$$|f(x) - P(x)| < \epsilon, \qquad |x| < N$$

befriedigt.

Dabei sind die Exponenten des Polinoms $P(x)$ Fouriersche Exponenten der Funktion $f(x)$, während die Koeffizienten von $P(x)$ durch Multiplikation von Fourierschen Koeffizienten der Funktion $f(x)$ mit gewissen Zahlen erhalten werden.

Beweis. Auf N-fastperiodische Funktionen überträgt sich leicht der Weylsche Beweis des Approximationssatzes der Theorie der fastperiodischen Funktionen.

Es seien $\epsilon > 0$ und $N > 0$ beliebige Zahlen. Da $f(x)$ N-fastperiodisch ist, so ist die Menge aller $(\epsilon/3, N)$ Verschiebungen der Funktion $f(x)$ fastperiodisch. Für diese Menge bilde man die Funktion

$$K_\delta(s) = \begin{cases} \dfrac{1}{\delta} & \text{für} \quad \tau_i \leqq s \leqq \tau_i + \delta \quad (i = 0, \pm 1, \cdots; \tau_{-i} = -\tau_i) \\ 0 & \text{für übrige } s \end{cases}$$

Wählt man $\delta(<1)$ so dass $|f(x+h) - f(x)| < \epsilon/3$ für $|h| < \delta$ und $|x| < N$, so hat man

$$\begin{aligned} \frac{1}{2n+1} \int_{-\tau_n}^{\tau_n+\delta} f(x+s)K_\delta(s)\, ds &= \frac{1}{2n+1} \sum_{j=-n}^{n} \frac{1}{\delta} \int_{\tau_j}^{\tau_j+\delta} f(x+s)\, ds \\ &= \frac{1}{2n+1} \sum_{j=-n}^{n} \frac{1}{\delta} \int_0^{\delta} f(x+\tau_j+s)\, ds. \end{aligned} \tag{2}$$

Nach der Eigenschaft der Zahlen τ_j und nach der Wahl von δ hat man

$$\left| \frac{1}{\delta} \int_0^\delta f(x + \tau_j + s)\, ds - f(x) \right| < \frac{2\epsilon}{3} \qquad (|x| < N).$$

Folglich

$$\left| \frac{1}{2n+1} \sum_{j=-n}^{n} \frac{1}{\delta} \int_0^\delta f(x + \tau_j + s)\, ds - f(x) \right| < \frac{2\epsilon}{3} \qquad (|x| < N). \tag{3}$$

Aus (2) und (3) folgt:

$$\left| \frac{1}{2n+1} \int_{-\tau_n}^{\tau_n+\delta} f(x + s) K_\delta(s)\, ds - f(x) \right| < \frac{2\epsilon}{3} \qquad (|x| < N). \tag{4}$$

Nun benützen wir die Fastperiodizität im Weylschen Sinne der Funktion $K_\delta(s)$—. Da $K_\delta(s)$ beschränkt ist, so gehört sie zur Klasse[5] W^2. Folglich kann man für jedes $\eta > 0$ solche Zahl $T_0 = T_0(\eta)$ und solche endliche trigonometrische Summe[6]

$$\sum_{k=1}^{g} a_k e^{-i\lambda_k s}$$

aufstellen, dass

$$\frac{1}{2T} \int_{-T}^{T} \left| K_\delta(s) - \sum_{k=1}^{g} a_k e^{-i\lambda_k s} \right|^2 ds < \eta$$

für $T > T_0$ ist.

Nach der Schwarzschen Ungleichung hat man:

$$\begin{aligned} &\left| \frac{1}{2n+1} \int_{-\tau_n}^{\tau_n+\delta} f(x+s) \left[K_\delta(s) - \sum_{k=1}^{g} a_k e^{-i\lambda_k s} \right] ds \right| \\ &\leqq \left[\frac{2\tau_n}{2n+1} \cdot \frac{1}{2\tau_n} \int_{-\tau_n}^{\tau_n+\delta} |f(x+s)|^2\, ds \right]^{\frac{1}{2}} \\ &\cdot \left[\frac{2\tau_n}{2n+1} \cdot \frac{1}{2\tau_n} \int_{-\tau_n}^{\tau_n+\delta} \left| K_\delta(s) - \sum_{k=1}^{g} a_k e^{-i\lambda_k s} \right|^2 ds \right]^{\frac{1}{2}}. \end{aligned} \tag{5}$$

Nun zeigen wir die Existenz des Limes

$$\lim_{n\to\infty} \frac{2\tau_n}{2n+1} = \lim_{n\to\infty} \frac{\tau_n}{n}$$

In der Tat, da $K_\delta(s)$, als eine im Weylschen Sinne fastperiodische Funktion, ein Mittelwert besitzt so ist:

$$\lim_{n\to\infty} \frac{1}{2\tau_n} \int_{-\tau_n}^{\tau_n} K_\delta(s)\, ds = \lim_{n\to\infty} \frac{2n+1}{2\tau_n}.$$

---

[5] Besicovitch, p. 77.

[6] Besicovitch, p. 91.

Bezeichnet man

$$M = \overline{\lim_{T\to\infty}} \frac{1}{2T} \int_{-T}^{T} |f(x)|^2 dx, \qquad m = \lim_{m\to\infty} \frac{\tau_n}{n}$$

und nimmt man

$$\eta < \frac{\epsilon^2}{9Mm^2}$$

so kommt man nach Berücksichtigung von

$$\lim_{n\to\infty} \frac{1}{2n+1} \int_{-\tau_n}^{\tau_n+\delta} f(x+s) \left(\sum_{k=1}^{g} a_k e^{-i\lambda_k s} \right) ds = \sum_{k=1}^{g} b_k e^{i\lambda_k x} = P(x)$$

wo

$$b_k = m \cdot a_k \cdot \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x) e^{-i\lambda_k x} dx$$

und (4), (5) zur Beziehung

$$|f(x) - P(x)| < \epsilon \qquad (|x| < N)$$

was zu beweisen war.

Aus dem Satze III ergibt sich unmittelbar der Eindeutigkeitssatz.

Satz IV. *Genügen zwei N-fastperiodischen Funktionen $f(x)$ und $g(x)$ der Ungleichung (I) und haben sie dieselbe Fouriersche Reihe, so sind sie identisch.*

Beweis. Nach dem Satze II ist die Differenz $h(x) = f(x) - g(x)$ eine N-fastperiodische Funktion; ihre Fouriersche Koefficienten sind alle gleich Null. Auf Grund des Approximationssatzes ist also $h(x) \equiv 0$.

3. Gehört eine N-fastperiodische Funktionen zur Klasse[7] B^2 so erfüllt sie die Ungleichung (I) und für sie existiert die Funktion $a(\lambda)$.

Mann kann auch eine notwendige und hinreichende Bedingung dafür angeben, dass für eine Funktion, die der Ungleichung (I) genügt, das Limes

$$a(\lambda) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x) e^{-i\lambda x} dx$$

existiere für alle reelle λ, so wie auch eine notwendige und hinreichende Bedingung für gleichmässige Existenz der mittelwerte (für jedes λ).[8]

Andererseits, gemeinsam mit B. Levine, habe ich konstruiert ein Beispiel einer beschränkten N-fastperiodischen Funktion $f(x)$, für welche der Mittelwert

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)\, dx$$

nicht mehr existiert.

---

[7] Besicovitch, p. 77 und p. 100.

[8] C. R. der Academie des U. S. S. R., vol. XIX, N 6–7, p. 447–450, Com. des Sc. Math. de Kharkoff, serie 4, t. XV_2, p. 3–35 (1938).

Dieses Beispiel wird in den C. R. USSR veröffentlicht.

Demgemäss erscheint es uns natürlich, eine Verallgemeinerung des Mittelwertbegriffes zu unternehmen. Am allgemeinesten lässt es sich durchführen, wenn man vom verallgemeinerten Limes von S. Banach Gebrauch macht.[9]

Sei $f(x)$ eine der Ungleichung (I) genügende Funktion. Alsdann ist $F(T) = \frac{1}{2T}\int_{-T}^{T} f(x)\,dx$ eine für alle T beschränkte Funktion. Man Bezeichne mit Lim das verallgemeinerte Limes von S. Banach. Unter den Mittelwert $M\{f(x)\}$ der Funktion $f(x)$ wollen wir verstehen $\mathrm{Lim}_{T\to\infty} F(T)$.

Aus den Eigenschaften des Verallgemeinerten Limes von S. Banach ergibt sich, dass der soeben definierte Mittelwert volgende Bedingungen erfüllt:

1) $M\{af + bg\} = am\{f\} + bm\{g\}$, ($a$ b sind konst., $f(x)$ und $g(x)$ genügen (I)).
2) $M\{f(x)\} \geqq 0$ wenn $f(x) \geqq 0$.
3) $M\{f(x + x_0)\} = M\{f(x)\}$ (x_0 ist reell).
4) $M\{1\} = 1$.
5) Existiert $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(x)\,dx$, so stimmt es mit $M\{f(x)\}$ überein.

Auf übliche Weise erhält man alsdann die Schwarzsche Ungleichung:

$$| M\{f(x)g(x)\} | \leqq [M\{| f(x) |^2\}\cdot M\{| g(x) |^2\}]^{\frac{1}{2}}$$

und bildet die Fouriersche Reihe

$$f(x) \sim \sum A_n e^{i\Lambda_n x}, \qquad A_n = M\{f(x)e^{-i\Lambda_n x}\}.$$

Es gelten für diese Fouriersche Reihen die Sätze III und IV. Wollen wir den Beweis z.B. für den Satz III durchführen.

Es sei auf der fastperiodischen Menge $\tau_i(\epsilon/3, N)$ $(i = 0, \pm 1, \pm 2, \cdots)$ die Funktion $K_\delta(s)$ wie folgt definiert:

$$K_\delta(s) = \begin{cases} \dfrac{1}{r\delta} & \text{für } \tau_i \leqq s \leqq \tau_i + \delta, \\ 0 & \text{für übrige } s, \end{cases}$$

wo

$$r = \lim_{n\to\infty} \frac{n}{\tau_n}.$$

Offenbar ist

$$M\{K_\delta(s)\} = \lim_{n\to\infty} \frac{1}{2\tau_n}\int_{-\tau_n}^{\tau_n} K_\delta(s)\,ds = \lim_{n\to\infty} \frac{2n}{2\tau_n}\cdot\frac{\delta}{\delta r} = 1.$$

Betrachten wir nun die Funktion

$$\varphi(x) = M_s\{f(x + s)K_\delta(s)\},$$

die existiert wegen (I) und der Schwarzschen Ungleichung.

---

[9] S. Banach, *Théorie des operations linéaires*, p. 33.

Wir haben

$$|f(x) - \varphi(x)| \leqq M_s\{|f(x+s) - f(x)| K_\delta(s)\}.$$

Ist aber $|x| < N$, so ist in denjenigen Intervallen wo $K_\delta(s) \neq 0$

$$|f(x+s) - f(x)| < \frac{2\epsilon}{3}$$

wenn nur δ genügend klein ist.

Somit hat man

$$|f(x) - \varphi(x)| < \frac{2\epsilon}{3} M_s\{K_\delta(s)\} = \frac{2\epsilon}{3}.$$

Nun approximiere man im Mittel die Funktion $K_\delta(s)$ durch endliche trigonometrische Summe und schliesse den Beweis des Satzes wie im Falle der Existenz der gewöhnlichen Mittelwerte.

4. Man kann eine andere Definition der N-fastperiodizität zu Grunde legen. Man betrachte die im Intervalle $-\infty < x < \infty$ stetige Funktionen $f(x)$, die den folgenden zwei Bedingungen genügen:

I. *Für jedes $\epsilon > 0$ und $N > 0$ existiert eine relativ dichte Menge reeller Zahlen $\tau = \tau(\epsilon, N)$ sodass*

$$|f(x \pm \tau) - f(x)| < \epsilon \quad (|x| < N).$$

II. $\tau(\epsilon, N) + \tau(\rho, N) = \tau(\delta; N)$, *wobei $\delta = \delta(\epsilon, \rho)$ strebt gegen 0 mit ϵ und ρ.*

Für diese Funktionenklasse gilt der Satz II (Invarianz gegenüber Addition und Multiplikation).

Setzt man noch voraus, dass $f(x)$ zur Klasse B^2 gehört, so bleibt auch der Approximations—und der Eindeutigkeitsatz bestehen.

Im allgemeinen, wenn man die Funktion $\delta(\epsilon, \rho)$ der Bedingung

$$\delta(\epsilon, \rho) = \epsilon + \lambda_\epsilon(\rho)$$

unterwirft wo $\lambda_\epsilon(\rho)$ mit ρ gegen 0 gleichmässig für alle ϵ strebt, so lässt sich zeigen, dass für jede $\epsilon > 0$ und $N > 0$ eine fastperiodische Menge (ϵ, N)-Verschiebungen existiert, sodass also $f(x)$ die in dieser Arbeit angenommene Bedingung erfüllt.

Ganz ausführlich behandelte ich alle diese Fragen anderswo.[10]

Kapitel 2. Über lineare Differentialgleichungen mit fastperiodischen Koeffizienten

1. In diesem Kapitel behandeln wir einige Eigenschaften beschränkter Lösungen der linearen Differentialgleichungen mit fastperiodischen Koeffi-

[10] Com. des. Sc. Math. de Kharkoff, series 4, t. XV_2, p. 3-35 (1938).

zienten. Wir wollen zeigen, dass unter gewissen Umständen solche Lösungen N-fastperiodische Funktionen sind.[11]

Wir betrachten ein System linearer Differentialgleichungen

$$(s_i)\quad \frac{dx_i}{dt} = f_{i,1}(t)x_1 + f_{i,2}(t)x_2 + \cdots + f_{i,n}(t)x_n + g_i(t) \qquad (i = 1, 2, \cdots n),$$

wo $f_{i,k}(t)$ und $g_i(t)$ sind stetig und reell für reelle t.

Für willkürlich gegebene Anfangsbedingungen, die durch Angabe der Werte $x_1(0, x_2(0), \cdots, x_n(0)$ festgesetzt sind, lässt sich (beispielsweise durch die Picardsche Methode der sukzessiven Approximation) eine Lösung finden, die diese Anfangswerte annimmt.

Nun betrachten wir eine Folge von Systemen:

$$(s_i^{(\rho)})\quad \frac{dx_i}{dt} = f_{i,1}^{(\rho)}(t)x_1 + f_{i,2}^{(\rho)}(t)x_2 + \cdots + f_{i,n}^{(\rho)}(t)x_n + g_i^{(\rho)}(t) \qquad (i = 1, 2, \cdots, n)$$

deren Koeffizienten konvergieren bzw. zu $f_{i,k}(t)$ und $g_i(t)$ gleichmässig in jeden endlichen Intervall, wenn $\rho \to \infty$ und betrachten wir solche Folgen von Lösungen dieser Systeme, für welche Anfangwerte $x_i^{(\rho)}(0) \to x_i(0)$ für $\rho \to \infty$; dann konvergieren die Lösungen $x_i^{(\rho)}(t)$ gleichmässig in jedem endlichen Intervall zur Lösung des Systems (s_i). Das ist eine Folgerung der Picardschen Methode.

Wählt man aus der Systemenfolge $(s_i^{(\rho)})$ zwei Systemen $(s_i^{(\rho_1)})$ und $(s_i^{(\rho_2)})$ so dass

$$\left.\begin{array}{ll} |f_{i,k}^{(\rho_1)}(t) - f_{i,k}^{(\rho_2)}(t)| < \epsilon & (i, k = 1, 2, \cdots, n) \\ |g_i^{(\rho_1)}(t) - g_i^{(\rho_2)}(t)| < \epsilon & (i = 1, 2, \cdots, n) \end{array}\right. \qquad (-\infty < t < \infty),$$

wo $\epsilon > 0$ willkürlich ist, so folgt aus der Picardschen Methode, dass wenn Anfandswerte der Lösungen $x_i^{(\rho_1)}(t)$ und $x_i^{(\rho_2)}(t)$ jeder Systemen den Ungleichungen

$$|x_i^{(\rho_1)}(0) - x_i^{(\rho_2)}(0)| < \alpha, \qquad (i = 1, 2, \cdots, n)$$

genügen, so lässt sich eine Zahl $\omega = \omega(\alpha, \epsilon)(>0)$ und eine Zahl $T(>0)$ angeben, für welche

$$(1)\qquad |x_i^{(\rho_1)}(t) - x_i^{(\rho_2)}(t)| < \omega,$$

für

$$(2)\qquad -T \leqq t \leqq T$$

ist, und, umgekehrt, für jedes noch so kleine $\omega > 0$ und so grosse T, lassen sich solche Zahlen α und ϵ angeben, dass die Ungleichungen (1) und (2) bestehen.

Nehmen wir nun an, dass im System (s_i) alle Funktionen $f_{i,k}(t)$ und $g_i(t)$ fastperiodisch sind. Es heisse dann das System fastperiodisch. Da die Anzahl der Funktionen $f_{i,k}(t)$ und $g_i(t)$ endlich ist, so existiert für jedes $\epsilon(>0)$ eine relativ dichte Menge $\{\tau(\epsilon)\}$ der ϵ-Verschiebungen, die allen Funktionen

[11] Was die Methoden und Bezeichnungen anbetrifft, die in diesem Kapitel auftreten, vergl. man die Abhandlung von J. Favard, *Sur les équations différentielles linéaire à coefficients presque périodiques*, Acta Math. t. 51, p. 31–81 (1927).

gemeinsam sind. Die Tatsache, dass $\tau(\epsilon)$ ϵ-Verschiebung für alle Funktionen $f_{i,k}(t)$ und $g_i(t)$ ist, bezeichnen wir wie folgt

$$|\,(s_{t+\tau}) - (s_t)\,| < \epsilon,$$

und es hiesse $\tau(\epsilon)$ eine Verschiebung für das System (s_t).

Mit $\bar{E}_\epsilon(s)$ bezeichnen wir ferner die Menge aller ganzen Verschiebungen des Systems (s_t). Bekanntlich[12] ist $\bar{E}_\epsilon(s)$ fastperiodische für fast alle ϵ.

Gleichzeitig mit dem System (s_t) wollen wir das homogene System

$$(\Sigma_t) \qquad \frac{dx_i}{dt} = f_{i,1}(t)x_1 + \cdots + f_{i,n}(t)x_n \qquad (i = 1, 2, \cdots, n)$$

betrachten.

Satz I. *Hat das inhomogene System (s_t) eine beschränkte, während das System (Σ_t) keine beschränkte Lösung (mit Ausnahme, selbst verständlich, der trivialen Lösung $x_i = 0$), so besteht die beschränkte Lösung von (s_t) aus N-fastperiodischen Funktionen.*

Beweis. Sei $\epsilon_1, \epsilon_2, \cdots$ eine unendliche Folge unbeschränkt abnehmender Zahlen, für die die Mengen $\bar{E}_{\epsilon_l}(s)$ fastperiodisch sind. Bezeichnet man die Zahlen der Menge $\bar{E}_{\epsilon_l}(s)$ mit τ_{lk} $(l = 1, 2, \cdots)$, $(k = 0, \pm 1, \pm 2, \cdots)$, so gilt offenbar

$$(3) \qquad |\,(s_{t+\tau_{lk}}) - (s_t)\,| < \epsilon_l \qquad (l = 1, 2, \cdots) \quad (k = 0, \pm 1, \pm 2, \cdots).$$

Ist nun $x_i(t)$ eine Lösung des Systems (s_t) so ist offenbar $x_i(t + \tau_{lk})$ eine Lösung des Systems $(s_{t+\tau_{lk}})$. Wegen früher gemachter Bemerkungen und (3) genügt es für den Beweis des Satzes I zu zeigen, dass

a) Die Zahlen $x_i(\tau_{lk})$ nur einen einzigen Grenzwert $x_i(0)$ für unbeschränkt wachende l und k haben;

b) für jedes i der Grenzwelt $x_i(\tau_{lk})$ für $l \to \infty$ gleichmässig nach k existiert, d.h. es lässt sich für jedes $\alpha > 0$ eine Zahl $N(\alpha)$ angeben, sodass für $l > N(\alpha)$ und für alle i und k

$$|\,x_i(\tau_{lk}) - x_i(0)\,| < \alpha$$

ist.

Fixieren wir k und lassen wir l unbegrenzt wachsen, so gilt wegen (3)

$$\lim_{l\to\infty} (s_{t+\tau_{lk}}) = (s_t).$$

Wenn wir annehmen, dass die Folge $x_i(\tau_{lk})$ für irgendein i und für fixierten k zwei Grenzwerte $x_i(0)$ und $x_i'(0)$ aufweise, so würde das homogene System eine

[12] Besicovitch, p. 55–59.

beschränkte Lösung zulassen, was der Bedingung des Satzes widerspricht. Auf dieselbe Weise schliessen wir, dass für kein i die Grenzwerte

$$\lim_{l\to\infty} x_i(\tau_{lk_1}), \qquad \lim_{l\to\infty} x_i(\tau_{lk_2}) \qquad (k_1 \neq k_2)$$

verschieden sein können. Somit ist a) bewiesen.

Zum Beweise von b) nehmen wir an, dass der Grenzwert $x_i(\tau_{lk})$ existiert ungleichmässig nach k. Dann lassen sich eine Zahl $\alpha > 0$ und zwei unendliche Folgen angeben

$$l_1, l_2, \cdots, l_p, \cdots$$

$$k_1, k_2, \ldots, k_p, \ldots$$

für welche bei gewissem i

$$(4) \qquad | x_i(\tau_{l_p k_p}) - x_i(0) | > \alpha$$

gilt.

Aus der unendlichen Folge $x_i(\tau_{l_p k_p})$ lässt sich aber eine konvergente Unterfolge auswählen, die etwa zu $x_i'(0)$ konvergiere. Wegen (4) gilt

$$| x_i'(0) - x_i(0) | > \alpha.$$

Dann wird has homogene System (Σ_t) eine beschränkte und von 0 verschiedene Lösung zulassen, was der Bedingung des Satzes wiederspricht. Somit ist der Satz I vollständig bewiesen.

2. Zum Beweise des Satzes I benötigten wir der Eindeutigkeit der Grenzwerte $x_i(\tau_{lk})$. J. Favard[13] aber hat gezeigt, dass sich diese Grenzwerte für gewisse Lösung $x_i(t)$ eindeutig bestimmen auch in dem Falle, wo das homogene System (Σ_t) keine Lösung hat, die nach dem absolutem Betrag noch so klein werden könnte (die Lösung $x_i = 0$ ausgenommen). Dabei verstehen wir unter dem absoluten Betrag den Ausdruck

$$[x_1^2(t) + x_2^2(t) + \cdots + x_n^2(t)]^{\frac{1}{2}}.$$

Wir haben also:

SATZ II. *Hat das System (Σ_t) keine lösung, derer absolute Betrag noch so klein werden könnte* (Lösung $x_i = 0$ ausgenommen) *und hat (s_t) beschränkte Lösungen, so besteht wenigstens eine dieser Lösungen aus N-fastperiodischen Funtionen.*

INSTITUT FÜR MATHEMATIK,
KHARKOFF, U. S. S. R.

[13] J. Favard, l. c., pp. 60–61.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

GRUNDZÜGE EINER INHALTSLEHRE IM HILBERTSCHEN RAUME

VON KARL LÖWNER

(Received February 21, 1939)

1. Einleitung

Bei dem üblichen axiomatischen Aufbau der Inhaltslehre im endlichdimensionalen euklidischen Raume wird das Lebesguesche Mass $\mu(\mathfrak{A})$ einer messbaren Punktmenge $\mathfrak{A}$ in seiner Abhängigkeit von $\mathfrak{A}$ als eine Mengenfunktion aufgefasst, welche folgenden Postulaten genügt:

1. *Der Definitionsbereich von $\mu(\mathfrak{A})$ ist ein σ-Körper* K, *welcher mit jeder Menge auch alle ihr kongruenten enthält.*[1]
2. *$\mu(\mathfrak{A})$ ist reell und nichtnegativ.*
3. *Ist $\mathfrak{A} = \sum_n \mathfrak{A}_n$,*[2] *$\mathfrak{A}_n <$ K$(n = 1, 2, \cdots)$, $\mathfrak{A}_k\mathfrak{A}_l = 0$ $(k \neq l)$ so ist*

$$\mu(\mathfrak{A}) = \sum_n \mu(\mathfrak{A}_n)$$

4. *Sind $\mathfrak{A}$ und $\mathfrak{B}$ kongruente Mengen aus* K, *so ist*

$$\mu(\mathfrak{A}) = \mu(\mathfrak{B})$$

5. *Jeder Würfel ist messbar und hat einen endlichen positiven Inhalt.*

Die vorliegende Arbeit enthält einen Versuch, eine entsprechende Theorie für den Hilbertschen Raum auszubauen.[3] Man sieht sofort, dass das Axiomensystem nicht wörtlich übernommen werden kann. Zunächst muss man im 4. Axiom die Würfel, welche im H.R.[4] nicht beschränkte Punktmengen darstellen, durch hier einfachere Gebilde, am besten durch *Kugeln* ersetzen. Doch auch bei dieser Modifikation ist das Axiomensystem nicht erfüllbar. Das sieht man so ein: Aus den drei ersten Axiomen folgt bekanntlich, dass $\mu(\mathfrak{A}) \leqq \mu(\mathfrak{B})$ ist, wenn $\mathfrak{A}$ in $\mathfrak{B}$ enthalten ist. Nun sei $\mathfrak{K}$ eine Kugel vom Radius $a > 0$. Man denke sich nun ein cartesisches Achsenkreuz des H.R., dessen Mittelpunkt mit

---

[1] Ein System von Mengen heisst bekanntlich ein Körper, wenn die Summen- und Differenzbildung aus ihm nicht herausführt. Es heisst ein σ-Körper, wenn sogar die Summenbildung mit abzählbar vielen Summanden aus dem System nicht herausführt. Der Ausdruck "Summe" wird hier wie stets im Folgenden gleichbedeutend mit "Vereinigung" gebraucht.—Kongruent sollen zwei Punktmenge des Hilbertschen Raumes heissen, wenn sie durch eine isometrische Abbildung des vollen Raumes auf sich selbst ineinander übergeführt werden können.

[2] Die Anzahl der Summanden kann endlich oder abzählbar unendlich sein.

[3] Einen kurzen Bericht über die vorliegende Untersuchung hat der Verfasser auf dem Kongress der Mathematiker der slavischen Länder, Prag 1934, vorgetragen. Siehe "Zprávy o druhém sjezdu matematiků zemí slovanských, Praha 1935."

[4] "H.R." ist Abkürzung für "Hilbertscher Raum."

dem Mittelpunkt von $\mathfrak{K}$ zusammenfällt. Um die Mittelpunkte der auf den positiven und negativen Achsen liegenden Radien von $\mathfrak{K}$ schlage man je eine Kugel vom Radius $b < a/(2\sqrt{2})$ $(b > 0)$. So erhält man eine Folge von Kugeln $\mathfrak{K}_n (n = 1, 2, \cdots)$ von gleichem Radius, welche paarweise punktfremd sind und alle in $\mathfrak{K}$ liegen. Bezeichnet man den Inhalt einer Kugel vom Radius c mit $I(c)$, so folgt aus dem dritten und vierten Postulat

$$\mu(\mathfrak{K}_1 + \mathfrak{K}_2 + \cdots + \mathfrak{K}_n) = nI(b)(n = 1, 2, \cdots)$$

Es ist also

$$nI(b) \leqq I(a) \ (n = 1, 2, \cdots)$$

Wegen Gültigkeit des Archimedischen Axioms im Bereich der positiven reellen Zahlen ist diese Ungleichung im Widerspruch zu dem modifizierten fünften Postulat.

Wir sehen, dass eine Abänderung des Axiomensystems notwendig ist. Gleichzeitig weist uns die eben durchgeführte Überlegung einen naturgemässen Weg hierzu: Man muss entweder den Bereich der nichtnegativen reellen Zahlen durch ein System von Grössen ersetzen, welches dem Archimedischen Axiom nicht genügt, oder man muss, wenn man den Bereich der reellen Zahlen nicht verlassen will, die Forderung der Nichtnegativität der Inhaltswerte aufgeben. *Wir werden hier den ersten Weg einschlagen.*

Der vorliegende Versuch einer Inhaltslehre in einem Raum von unendlich vielen Dimensionen unterscheidet sich von den bisherigen wesentlich darin, dass diese alle den Bereich der Inhaltswerte dem System der nichtnegativen reellen Zahlen entnehmen. Der Weg, den wir einschlagen, ist kein rein axiomatischer. Wir kombinieren die axiomatische Methode mit geometrischen Betrachtungen, indem wir den Inhalt zunächst nur für gewisse Rotationskörper definieren, welche als "Verwandte" der Kugel hier eine ähnliche Rolle spielen wie die Polyeder im endlichdimensionalen Raum. Indem wir das Cavalierische Prinzip als heuristisches Prinzip benutzen, werden wir auf naturgemässe Begriffsbildungen geführt.

2. Geometrische Sätze über Rotations—und Rotativkörper

Unter einem Rotationskörper $\mathfrak{R}$ des H. R. soll im Folgenden ein Körper verstanden werden, welcher alle Rotationen um einen endlichdimensionalen ebenen Unterraum $\mathfrak{a}$[5] des H. R. gestattet. Unter einer Rotation um $\mathfrak{a}$ ist jede isometrische Abbildung des H. R. in sich zu verstehen, welche alle Punkte von $\mathfrak{a}$ einzeln festlässt.[6] $\mathfrak{a}$ nennen wir eine *Achse* von $\mathfrak{R}$.

Um spätere Betrachtungen nicht unterbrechen zu müssen, wollen wir hier

[5] Zu den ebenen Räumen zählen wir auch einzelne Punkte. Wir schreiben ihnen die Dimension Null zu.

[6] Wenn man diese Definition einer Rotation bei euklidischen Räumen endlicher Dimensionen verwendet, muss man etwa eine Spiegelung an einer Ebene auch als Rotation bezeichnen.

zunächst eine Reihe von geometrischen Hilfssätzen über Rotationskörper zusammenstellen.

1. Hilfssatz: *Ist $\mathfrak{a}_k$ eine k-dimensionale Achse des Rotationskorpers $\mathfrak{R}$, so ist auch jeder endlichdimensionale ebene Raum $\mathfrak{a}_l$ von einer Dimension $l > k$, welcher $\mathfrak{a}_k$ enthält, ebenfalls Achse von $\mathfrak{R}$.*

Dieser Satz folgt unmittelbar aus der Definition einer Rotationsachse.

2. Hilfssatz: *Die Rotationskörper bilden einer Mengenkörper.* Beweis: Der Rotationskörper $\mathfrak{R}$ habe die Achse $\mathfrak{a}$, der Rotationskörper $\mathfrak{S}$ die Achse $\mathfrak{b}$. Man bilde den Verbindungsraum $\mathfrak{c}$ von $\mathfrak{a}$ und $\mathfrak{b}$, d. h. den ebenen Raum kleinster Dimension, welcher sowohl $\mathfrak{a}$ als $\mathfrak{b}$ enthält. Seine Dimension ist bekanntlich nicht grösser als die Summe der Dimensionen von $\mathfrak{a}$ und $\mathfrak{b}$ vermehrt um eins. Nach dem l. Hilfssatz ist $\mathfrak{c}$ Achse sowohl von $\mathfrak{R}$ als auch von $\mathfrak{S}$. Hieraus folgt offenbar, dass auch $\mathfrak{R} + \mathfrak{S}$ und wenn $\mathfrak{R} < \mathfrak{S}$ ist, auch $\mathfrak{S} - \mathfrak{R}$ sämtliche Rotationen um $\mathfrak{c}$ gestattet.

3. Hilfssatz: *Sind $\mathfrak{a}$ und $\mathfrak{b}$ Achsen des Rotationskörpers $\mathfrak{R}$, so ist auch der Verbindungsraum von $\mathfrak{a}$ und $\mathfrak{b}$ Achse von $\mathfrak{R}$.*

Der Satz ist eine unmittelbare Folge des 1. Hilfssatzes.

4. Hilfssatz: *Sind $\mathfrak{a}$ und $\mathfrak{b}$ zwei Achsen des Rotationskörpers $\mathfrak{R}$ und haben $\mathfrak{a}$ und $\mathfrak{b}$ einen nichtleeren Durchschnitt $\mathfrak{d}$, dann ist auch dieser Achse von $\mathfrak{R}$.*

Wir beweisen die etwas *schärfere Aussage*:

4′. Hilfssatz: *Jede Rotation um den nichtleeren Durchschnitt $\mathfrak{d}$ zweier ebener Räume endlicher Dimension $\mathfrak{a}$ und $\mathfrak{b}$, lässt sich aus endlich vielen Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ zusammensetzen.*

Wir schliessen durch vollständige Induktion nach dem Wert von $s = p + q$ worin p und q die Dimensionen von $\mathfrak{a}$ und $\mathfrak{b}$ bedeuten. Da man annehmen kann, dass keiner der beiden Räume $\mathfrak{a}$ und $\mathfrak{b}$ in dem anderen ganz enthalten ist, ist der kleinstmögliche Wert von s gleich 2 und wird für $p = q = 1$ geliefert. $\mathfrak{a}$ und $\mathfrak{b}$ sind dann zwei sich in einem Punkt D schneidende Geraden. Hier kann man folgendermassen schliessen. Durch eine Rotation B um $\mathfrak{b}$ entstehe aus $\mathfrak{a}$ die Gerade $\mathfrak{a}^* \neq \mathfrak{a}$. Sie geht ebenfalls durch D hindurch. Jede Rotation A^* um $\mathfrak{a}^*$ lässt sich vermöge einer zugehörigen Rotation A um $\mathfrak{a}$ in der Form $\mathrm{A}^* = \mathrm{B}^{-1}\mathrm{AB}$ darstellen. Wenn also der zu beweisende Satz für das Achsenpaar $\mathfrak{a}, \mathfrak{a}^*$ richtig ist, so auch für das Paar $\mathfrak{a}, \mathfrak{b}$. Sei nun δ der zwischen 0 und $\frac{1}{2}\pi$ gelegene Winkel zwischen $\mathfrak{a}$ und $\mathfrak{b}$. Der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ variiert mit $\mathfrak{a}^*$ zwischen 0 und 2δ. Ist $2\delta \geqq \frac{1}{2}\pi$, so kann man $\mathfrak{a}^*$ so wählen, dass der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ gleich $\frac{1}{2}\pi$ ist. Ist aber $2\delta < \frac{1}{2}\pi$, so wähle man $\mathfrak{a}^*$ so, dass der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ den maximalen Wert 2δ annimmt und verfahre dann mit $\mathfrak{a}$ und $\mathfrak{a}^*$ ebenso wie vorher mit $\mathfrak{a}$ und $\mathfrak{b}$. Nach endlich vielen Schritten dieser Art gelangt man schliesslich zu einem Achsenpaar $\mathfrak{a}', \mathfrak{b}'$ durch D, welche aufeinander senkrecht stehen und die wesentliche Eigenschaft haben, dass jede Rotation um $\mathfrak{a}'$ bzw. $\mathfrak{b}'$ sich aus endlich vielen Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ zusammensetzen lässt. Ebenso wie im dreidimensionalen euklidischen Raume beweist man aber, dass jede Rotation um den Schnittpunkt zweier einander

senkrecht schneidender Geraden $\mathfrak{a}'$, $\mathfrak{b}'$ aus höchstens *drei* Rotationen um $\mathfrak{a}'$ und $\mathfrak{b}'$ zusammengesetzt werden kann.

Sei nun $p + q > 2$. Ist die Dimension d des Durchschnitts $\mathfrak{d}$ von $\mathfrak{a}$ und $\mathfrak{b}$ *positiv*, so ist jede Rotation um $\mathfrak{d}$ (und um so mehr jede Rotation um $\mathfrak{a}$ bzw. $\mathfrak{b}$) vollkommen bestimmt durch ihr Verhalten in einem vollständigen Orthogonalraum[7] $\mathfrak{O}$ zu $\mathfrak{d}$. Die Schnitte von $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{d}$ mit $\mathfrak{O}$ haben der Reihe nach die Dimensionen $p' = p - d$, $q' = q - d$, $d' = 0$. Die Summe $s' = p' + q' = p + q - 2d$ ist *kleiner* als $s = p + q$. Man kann also vollständige Induktion anwenden.

Sei jetzt $d = 0, p > 0, q > 0, p + q > 2$. Der Durchschnitt ist ein einzelner Punkt D. Ohne Beschränkung der Allgemeinheit kann angenommen werden, dass $p > 1$ ist. Man verstehe unter $\mathfrak{a}'$ irgend einen durch D gehenden Unterraum von $\mathfrak{a}$ von der Dimension $p - 1$. Durch eine passende Rotation um $\mathfrak{b}$ kann $\mathfrak{a}$ in eine Lage $\mathfrak{a}^*$ gebracht werden derart, dass $\mathfrak{a}$ und $\mathfrak{a}^*$ genau $\mathfrak{a}'$ zum Durchschnitt haben. Die Dimensionen von $\mathfrak{a}$ und $\mathfrak{a}^*$ sind beide gleich p und die ihrer Schnitte mit einem vollständigen Orthogonalraum zu $\mathfrak{a}'$ gleich 1. Da die Summe der letzteren gleich 2 ist, kann nach den früheren Überlegungen behauptet werden, dass jede Rotation um $\mathfrak{a}'$ aus Rotationen um $\mathfrak{a}$ und $\mathfrak{a}^*$ zusammengesetzt werden kann. Ebenso wie im Falle $p = q = 1$ ist jede Rotation um $\mathfrak{a}^*$ durch Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ ausdrückbar. Im Ganzen kann man jede Rotation um $\mathfrak{a}'$ aus Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ zusammensetzen. Nun betrachte man das Achsenpaar $\mathfrak{a}'$, $\mathfrak{b}$. Es hat eine um 1 kleinere Dimensionssumme als das Paar $\mathfrak{a}$, $\mathfrak{b}$. Man kann also wieder vollständige Induktion anwenden. Da der Schnitt von $\mathfrak{a}'$ und $\mathfrak{b}$ gleich dem Punkt D ist, ist der Satz somit vollständig bewiesen.

Es erweist sich notwendig, die bisher aufgestellten Sätze noch ein wenig zu *verschärfen*. Der H. R. kann ebenso wie der endlich dimensionale euklidische Raum zu einem *projektiven Raum* ergänzt werden, indem man jeder Richtung einen unendlich fernen Punkt der Geraden zuordnet, welche die gegebene Richtung besitzt. Man kann dann auch in der üblichen Weise unendlich ferne ebene Räume einführen. Ferner kann man den Begriff der Orthogonalität zweier Räume auf den Fall ausdehnen, dass einer von ihnen oder beide ins Unendliche fallen; denn schon bei eigentlichen Räumen ist ja die Aussage der Orthogonalität eine Aussage über ihre Schnitte mit dem Unendlichfernen.

Unter einer *Rotation um eine unendlich ferne endlichdimensionale Achse* $\mathfrak{a}$ verstehen wir jede isometrische Abbildung des H. R. in sich, welche *jeden Punkt von* $\mathfrak{a}$ *einzeln festhält und jeden vollständigen Orthogonalraum zu* $\mathfrak{a}$ *in sich überführt.*

Zur formalen Abrundung führen noch wir den leeren Raum $\mathfrak{d} = 0$ ein, dem wir die Dimension -1 zuschreiben. *Unter einer Rotation um* $\mathfrak{d} = 0$ *kann jede isometrische Abbildung des H. R. in sich verstanden werden.* Die einzigen Rotationskörper mit $\mathfrak{d} = 0$ als Achse sind der ganze H. R. und die Nullmenge.

Nach Einführung dieser Begriffserweiterungen können wir die Hilfssätze, die

---

[7] Ein *vollständiger Orthogonalraum* zu $\mathfrak{a}$ spannt mit $\mathfrak{a}$ den ganzen H.R. auf.

zunächst eine Reihe von geometrischen Hilfssätzen über Rotationskörper zusammenstellen.

1. HILFSSATZ: *Ist* $\mathfrak{a}_k$ *eine k-dimensionale Achse des Rotationskorpers* $\mathfrak{R}$, *so ist auch jeder endlichdimensionale ebene Raum* $\mathfrak{a}_l$ *von einer Dimension* $l > k$, *welcher* $\mathfrak{a}_k$ *enthält, ebenfalls Achse von* $\mathfrak{R}$.

Dieser Satz folgt unmittelbar aus der Definition einer Rotationsachse.

2. HILFSSATZ: *Die Rotationskörper bilden einer Mengenkörper.* BEWEIS: Der Rotationskörper $\mathfrak{R}$ habe die Achse $\mathfrak{a}$, der Rotationskörper $\mathfrak{S}$ die Achse $\mathfrak{b}$. Man bilde den Verbindungsraum $\mathfrak{c}$ von $\mathfrak{a}$ und $\mathfrak{b}$, d. h. den ebenen Raum kleinster Dimension, welcher sowohl $\mathfrak{a}$ als $\mathfrak{b}$ enthält. Seine Dimension ist bekanntlich nicht grösser als die Summe der Dimensionen von $\mathfrak{a}$ und $\mathfrak{b}$ vermehrt um eins. Nach dem l. Hilfssatz ist $\mathfrak{c}$ Achse sowohl von $\mathfrak{R}$ als auch von $\mathfrak{S}$. Hieraus folgt offenbar, dass auch $\mathfrak{R} + \mathfrak{S}$ und wenn $\mathfrak{R} < \mathfrak{S}$ ist, auch $\mathfrak{S} - \mathfrak{R}$ sämtliche Rotationen um $\mathfrak{c}$ gestattet.

3. HILFSSATZ: *Sind* $\mathfrak{a}$ *und* $\mathfrak{b}$ *Achsen des Rotationskörpers* $\mathfrak{R}$, *so ist auch der Verbindungsraum von* $\mathfrak{a}$ *und* $\mathfrak{b}$ *Achse von* $\mathfrak{R}$.

Der Satz ist eine unmittelbare Folge des 1. Hilfssatzes.

4. HILFSSATZ: *Sind* $\mathfrak{a}$ *und* $\mathfrak{b}$ *zwei Achsen des Rotationskörpers* $\mathfrak{R}$ *und haben* $\mathfrak{a}$ *und* $\mathfrak{b}$ *einen nichtleeren Durchschnitt* $\mathfrak{d}$, *dann ist auch dieser Achse von* $\mathfrak{R}$.

Wir beweisen die etwas *schärfere Aussage*:

4′. HILFSSATZ: *Jede Rotation um den nichtleeren Durchschnitt* $\mathfrak{d}$ *zweier ebener Räume endlicher Dimension* $\mathfrak{a}$ *und* $\mathfrak{b}$, *lässt sich aus endlich vielen Rotationen um* $\mathfrak{a}$ *und* $\mathfrak{b}$ *zusammensetzen.*

Wir schliessen durch vollständige Induktion nach dem Wert von $s = p + q$ worin p und q die Dimensionen von $\mathfrak{a}$ und $\mathfrak{b}$ bedeuten. Da man annehmen kann, dass keiner der beiden Räume $\mathfrak{a}$ und $\mathfrak{b}$ in dem anderen ganz enthalten ist, ist der kleinstmögliche Wert von s gleich 2 und wird für $p = q = 1$ geliefert. $\mathfrak{a}$ und $\mathfrak{b}$ sind dann zwei sich in einem Punkt D schneidende Geraden. Hier kann man folgendermassen schliessen. Durch eine Rotation B um $\mathfrak{b}$ entstehe aus $\mathfrak{a}$ die Gerade $\mathfrak{a}^* \neq \mathfrak{a}$. Sie geht ebenfalls durch D hindurch. Jede Rotation A^* um $\mathfrak{a}^*$ lässt sich vermöge einer zugehörigen Rotation A um $\mathfrak{a}$ in der Form $\mathrm{A}^* = \mathrm{B}^{-1}\mathrm{A}\mathrm{B}$ darstellen. Wenn also der zu beweisende Satz für das Achsenpaar $\mathfrak{a}, \mathfrak{a}^*$ richtig ist, so auch für das Paar $\mathfrak{a}, \mathfrak{b}$. Sei nun δ der zwischen 0 und $\frac{1}{2}\pi$ gelegene Winkel zwischen $\mathfrak{a}$ und $\mathfrak{b}$. Der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ variiert mit $\mathfrak{a}^*$ zwischen 0 und 2δ. Ist $2\delta \geqq \frac{1}{2}\pi$, so kann man $\mathfrak{a}^*$ so wählen, dass der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ gleich $\frac{1}{2}\pi$ ist. Ist aber $2\delta < \frac{1}{2}\pi$, so wähle man $\mathfrak{a}^*$ so, dass der Winkel zwischen $\mathfrak{a}$ und $\mathfrak{a}^*$ den maximalen Wert 2δ annimmt und verfahre dann mit $\mathfrak{a}$ und $\mathfrak{a}^*$ ebenso wie vorher mit $\mathfrak{a}$ und $\mathfrak{b}$. Nach endlich vielen Schritten dieser Art gelangt man schliesslich zu einem Achsenpaar $\mathfrak{a}'$, $\mathfrak{b}'$ durch D, welche aufeinander senkrecht stehen und die wesentliche Eigenschaft haben, dass jede Rotation um $\mathfrak{a}'$ bzw. $\mathfrak{b}'$ sich aus endlich vielen Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ zusammensetzen lässt. Ebenso wie im dreidimensionalen euklidischen Raume beweist man aber, dass jede Rotation um den Schnittpunkt zweier einander

senkrecht schneidender Geraden $\mathfrak{a}'$, $\mathfrak{b}'$ aus höchstens *drei* Rotationen um $\mathfrak{a}'$ und $\mathfrak{b}'$ zusammengesetzt werden kann.

Sei nun $p + q > 2$. Ist die Dimension d des Durchschnitts $\mathfrak{d}$ von $\mathfrak{a}$ und $\mathfrak{b}$ *positiv*, so ist jede Rotation um $\mathfrak{d}$ (und um so mehr jede Rotation um $\mathfrak{a}$ bzw. $\mathfrak{b}$) vollkommen bestimmt durch ihr Verhalten in einem vollständigen Orthogonalraum[7] $\mathfrak{O}$ zu $\mathfrak{d}$. Die Schnitte von $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{d}$ mit $\mathfrak{O}$ haben der Reihe nach die Dimensionen $p' = p - d$, $q' = q - d$, $d' = 0$. Die Summe $s' = p' + q' = p + q - 2d$ ist *kleiner* als $s = p + q$. Man kann also vollständige Induktion anwenden.

Sei jetzt $d = 0, p > 0, q > 0, p + q > 2$. Der Durchschnitt ist ein einzelner Punkt D. Ohne Beschränkung der Allgemeinheit kann angenommen werden, dass $p > 1$ ist. Man verstehe unter $\mathfrak{a}'$ irgend einen durch D gehenden Unterraum von $\mathfrak{a}$ von der Dimension $p - 1$. Durch eine passende Rotation um $\mathfrak{b}$ kann $\mathfrak{a}$ in eine Lage $\mathfrak{a}^*$ gebracht werden derart, dass $\mathfrak{a}$ und $\mathfrak{a}^*$ genau $\mathfrak{a}'$ zum Durchschnitt haben. Die Dimensionen von $\mathfrak{a}$ und $\mathfrak{a}^*$ sind beide gleich p und die ihrer Schnitte mit einem vollständigen Orthogonalraum zu $\mathfrak{a}'$ gleich 1. Da die Summe der letzteren gleich 2 ist, kann nach den früheren Überlegungen behauptet werden, dass jede Rotation um $\mathfrak{a}'$ aus Rotationen um $\mathfrak{a}$ und $\mathfrak{a}^*$ zusammengesetzt werden kann. Ebenso wie im Falle $p = q = 1$ ist jede Rotation um $\mathfrak{a}^*$ durch Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ ausdrückbar. Im Ganzen kann man jede Rotation um $\mathfrak{a}'$ aus Rotationen um $\mathfrak{a}$ und $\mathfrak{b}$ zusammensetzen. Nun betrachte man das Achsenpaar $\mathfrak{a}'$, $\mathfrak{b}$. Es hat eine um 1 kleinere Dimensionssumme als das Paar $\mathfrak{a}$, $\mathfrak{b}$. Man kann also wieder vollständige Induktion anwenden. Da der Schnitt von $\mathfrak{a}'$ und $\mathfrak{b}$ gleich dem Punkt D ist, ist der Satz somit vollständig bewiesen.

Es erweist sich notwendig, die bisher aufgestellten Sätze noch ein wenig zu *verschärfen*. Der H. R. kann ebenso wie der endlich dimensionale euklidische Raum zu einem ***projektiven Raum*** ergänzt werden, indem man jeder Richtung einen unendlich fernen Punkt der Geraden zuordnet, welche die gegebene Richtung besitzt. Man kann dann auch in der üblichen Weise unendlich ferne ebene Räume einführen. Ferner kann man den Begriff der Orthogonalität zweier Räume auf den Fall ausdehnen, dass einer von ihnen oder beide ins Unendliche fallen; denn schon bei eigentlichen Räumen ist ja die Aussage der Orthogonalität eine Aussage über ihre Schnitte mit dem Unendlichfernen.

Unter einer ***Rotation um eine unendlich ferne endlichdimensionale Achse*** $\mathfrak{a}$ verstehen wir jede isometrische Abbildung des H. R. in sich, welche ***jeden Punkt von*** $\mathfrak{a}$ ***einzeln festhält und jeden vollständigen Orthogonalraum zu*** $\mathfrak{a}$ ***in sich überführt.***

Zur formalen Abrundung führen noch wir den leeren Raum $\mathfrak{d} = 0$ ein, dem wir die Dimension $- 1$ zuschreiben. *Unter einer Rotation um* $\mathfrak{d} = 0$ *kann jede isometrische Abbildung des H. R. in sich verstanden werden.* Die einzigen Rotationskörper mit $\mathfrak{d} = 0$ als Achse sind der ganze H. R. und die Nullmenge.

Nach Einführung dieser Begriffserweiterungen können wir die Hilfssätze, die

[7] Ein *vollständiger Orthogonalraum* zu $\mathfrak{a}$ spannt mit $\mathfrak{a}$ den ganzen H.R. auf.

bisher aufgestellt worden sind, in dem Sinne *verschärfen, dass sie ihre Richtigkeit behalten auch in dem Falle, wo einer oder mehrere der darin auftretenden Räume unendlich fern sind oder den leeren Raum darstellen.* Die einfachen Beweise für diese Erweiterungen, die man übrigens ein Falle $\mathfrak{b} \neq 0$ als Grenzfälle der bisherigen Aussagen betrachten kann, können dem Leser überlassen bleiben. *Wenn die Hilfssätze im Folgendem zitiert werden, sind sie stets in der verschärften Form gemeint.*

Bemerkung: Beim Beweise der Hilfssätze ist von der Unendlichdimensionalität des H. R. kein Gebrauch gemacht worden. Die Hilfssätze 1, 2, 3 und ihre Beweise gelten wörtlich in endlichdimensionalen euklidischen Raum. Beim Beweise der Hilfssätze 4 und 4′ ist davon Gebrauch gemacht worden, dass der Verbindungsraum von $\mathfrak{a}$ und $\mathfrak{b}$ nicht der Vollraum ist. Unter dieser Voraussetzung gelten diese Sätze also auch in euklidischen Räumen endlicher Dimension.

Wir sind jetzt im Stande, eine Ubersichtüber *die Gesamtheit aller Achsen eines Rotationskörpers* $\mathfrak{R}$[8] zu gewinnen. Es sei $\mathfrak{a}$ eine Achse von $\mathfrak{R}$ *kleinstmöglicher* Dimension. *Ich behaupte, dass es nur eine Achse dieser Art gibt.* Gäbe es nämlich zwei verschiedene, $\mathfrak{a}$, $\mathfrak{b}$ so müsste ihr Durchschnitt nach dem (in der verschärften Form angewandten) 4. Hilfssatz auch Achse sein und diese hätte eine kleinere Dimension als $\mathfrak{a}$ und $\mathfrak{b}$. Wenn im Folgenden von *der* Achse $\mathfrak{a}_{\mathfrak{R}}$ von $\mathfrak{R}$ die Rede ist, so ist stets diese Achse kleinster Dimension gemeint. Nun sei $\mathfrak{a}$ eine von $\mathfrak{a}_{\mathfrak{R}}$ verschiedene Achse. Da der Durchschnitt von $\mathfrak{a}$ und $\mathfrak{a}_{\mathfrak{R}}$ keine kleinere Dimension haben darf als $\mathfrak{a}_{\mathfrak{R}}$, muss $\mathfrak{a}$ die Achse $\mathfrak{a}_{\mathfrak{R}}$ *enthalten.* Zusammenfassend können wir also behaupten:

5. Hilfssatz: *Unter den Achsen eines Rotationskörpers gibt es genau eine,* $\mathfrak{a}_{\mathfrak{R}}$, *kleinster Dimension. Man bekommt alle übrigen, indem man alle möglichen ebenen Räume endlicher Dimension durch* $\mathfrak{a}_{\mathfrak{R}}$ legt.

Von nun an sollen nur *beschränkte* Rotationskörper $\mathfrak{R}$ betrachtet werden. Die Achse $\mathfrak{a}_{\mathfrak{R}}$ eines solchen ist offenbar im *Endlichen* gelegen.

Eine wesentliche Rolle spielt in späteren Überlegungen eine Grösse, die jedem beschränkten Rotationskörper $\mathfrak{R}$ zugeordnet werden kann und die wir als *Radius* $\rho_{\mathfrak{R}}$ von $\mathfrak{R}$ bezeichnen. Wir verstehen darunter die *obere Grenze der Distanzen der Punkte von* $\mathfrak{R}$ *von der Achse* $\mathfrak{a}_{\mathfrak{R}}$. Wir beweisen jetzt eine Reihe von wichtigen Aussagen über $\rho_{\mathfrak{R}}$.

6. Hilfssatz: *Sei* $\mathfrak{a}$ *irgend eine Achse des Rotationskörper* $\mathfrak{R}$. *Dann ist die obere Grenze der Distanzen der Punkte von* $\mathfrak{R}$ *von* $\mathfrak{a}$ *wieder gleich* $\rho_{\mathfrak{R}}$.

Der Beweis folgt fast unmittelbar aus der Definition von $\rho_{\mathfrak{R}}$ unter Zuhilfnahme des 5. Hilfssatzes.

7. Hilfssatz: *Für irgendwelche Rotationskörper* $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_m$ *gilt die Formel*

$$\rho_{\mathfrak{R}_1+\mathfrak{R}_2+\cdots+\mathfrak{R}_m} = \text{Max.}\,(\rho_{\mathfrak{R}_1}, \rho_{\mathfrak{R}_2}, \cdots, \rho_{\mathfrak{R}_m}). \tag{1'}$$

[8] $\mathfrak{R}$ sei weder der Vollraum noch die Nullmenge.

Es genügt, den Satz für $m = 2$ zu beweisen. Man verstehe unter $\mathfrak{v}$ den Verbindungsraum von $\mathfrak{a}_{\mathfrak{R}_1}$ und $\mathfrak{a}_{\mathfrak{R}_2}$. Er ist Achse sowohl von $\mathfrak{R}_1$ als auch von $\mathfrak{R}_2$, also auch von $\mathfrak{R}_1 + \mathfrak{R}_2$. Die Grössen $\rho_{\mathfrak{R}_1}$, $\rho_{\mathfrak{R}_2}$, $\rho_{\mathfrak{R}_1+\mathfrak{R}_2}$ sind nach dem 6. Hilfssatz der Reihe nach gleich den oberen Grenzen der Distanzen der Punkte von $\mathfrak{R}_1$, $\mathfrak{R}_2$, $\mathfrak{R}_1 + \mathfrak{R}_2$ von $\mathfrak{v}$. Hieraus ersieht man unmittelbar die Richtigkeit des Satzes.

Im Hinblick auf unsere spätere Absicht, Rotationskörper zur *Approximation* allgemeiner Punktmengen zu benützen, kann die Frage aufgeworfen werden, ob der letzte Hilfssatz auf den Fall unendlich vieler Summanden ausgedehnt werden kann. Dies ist mit einer gewissen *Einschränkung* richtig. Es gilt nämlich der wichtige

8. Hilfssatz: $\mathfrak{R}$; $\mathfrak{R}_1$, $\mathfrak{R}_2$, $\cdots$ *seien lauter Rotationskörper und*

$$\mathfrak{R} = \mathfrak{R}_1 + \mathfrak{R}_2 + \cdots. \tag{2}$$

Ferner sei

$$\lim_{n\to\infty} \rho_{\mathfrak{R}_n} = 0.^{9} \tag{3}$$

Dann ist

$$\rho_{\mathfrak{R}} = \text{Max.}\,(\rho_{\mathfrak{R}_1}, \rho_{\mathfrak{R}_2}, \cdots). \tag{1}$$

Beweis: Da $\mathfrak{R}_n$ in $\mathfrak{R}$ enthalten ist, hat man

$$\rho_{\mathfrak{R}_n} \leqq \rho_{\mathfrak{R}} \qquad (n = 1, 2, \cdots), \tag{4}$$

also

$$\sigma_0 = \text{Max.}\,(\rho_{\mathfrak{R}_1}, \rho_{\mathfrak{R}_2}, \cdots) \leqq \rho_{\mathfrak{R}}. \tag{5}$$

Wir haben nachzuweisen, dass *in* (5) *das Gleichheitszeichen gültig ist.* Wir betrachten zunächst den Spezialfall, dass $\mathfrak{R}$ eine abgeschlossene Kugel $\mathfrak{K}$ ist. Wir schliessen *indirekt*. Sei $\sigma_0 < \rho_{\mathfrak{R}}$. Auf Grund des 7. Hilfssatzes ist für jedes $\mathfrak{S}_n = \mathfrak{R}_1 + \mathfrak{R}_2 + \cdots + \mathfrak{R}_n$ $(n = 1, 2, \cdots)$

$$\rho_{\mathfrak{S}_n} \leqq \sigma_0 \qquad (n = 1, 2, \cdots). \tag{6}$$

Man wähle den Index $n = n_1$ so, dass

$$\sigma_1 = \text{Max.}\,(\rho_{\mathfrak{R}_{n_1+1}}, \rho_{\mathfrak{R}_{n_1+2}}, \cdots) < \frac{\rho_{\mathfrak{R}} - \sigma_0}{4} = \rho_1 \tag{7}$$

ausfällt. Dies ist wegen $\lim_{n\to\infty} \rho_{\mathfrak{R}_n} = 0$ möglich. Aus der Definition von $\rho_{\mathfrak{S}_{n_1}}$ und (6) ergibt sich offenbar, dass man innerhalb $\mathfrak{K}$ eine abgeschlossene Kugel $\mathfrak{K}_1$ vom Radius ρ_1 wählen kann, welche keinen Punkt von $\mathfrak{S}_{n_1}$ enthält. Die Ungleichung (7) lehrt nun, dass man nach Streichung von $\mathfrak{R}_1$, $\mathfrak{R}_2$, $\cdots$, $\mathfrak{R}_{n_1}$ diesen Prozess wiederholen kann. Man erhält so eine Reihe von Indizes

$$n_1 < n_2 < \cdots,$$

[9] Ohne diese Zusatzforderung ist der Satz unrichtig. Man kann z. B. die Einheitskugel mit abzählbar vielen Kugeln vom gleichen Radius $\rho < 1$ überdecken.

und eine Reihe von Kugeln

$$\mathfrak{K} > \mathfrak{K}_1 > \mathfrak{K}_2 > \cdots$$

derart, dass $\mathfrak{K}_i$ keinen Punkt von $\mathfrak{S}_{n_i}$ $(i = 1, 2, \cdots)$ enthält. Nun gibt es aber einen Punkt T, der in allen Kugeln $\mathfrak{K}_i$ enthalten ist. Dieser ist wegen der Abgeschlossenheit von $\mathfrak{K}$ in $\mathfrak{K}$ enthalten, aber in keinem $\mathfrak{R}_n$ $(n = 1, 2, \cdots)$. Dies steht mit $\mathfrak{K} = \mathfrak{R}_1 + \mathfrak{R}_2 + \cdots$ im Widerspruch.

Wir gehen nun zu einem allgemeinen Rotationskörper $\mathfrak{R}$ über. Wir nehmen wieder an, dass $\sigma_0 < \rho_{\mathfrak{R}}$ ist. Dann existiert ein Punkt T in $\mathfrak{R}$, dessen Distanz von $\mathfrak{a}_{\mathfrak{R}}$ *grösser als* σ_0 ist. Durch T lege man den vollständigen Orthogonalraum $\mathfrak{h}'$ zu $\mathfrak{a}_{\mathfrak{R}}$. $\mathfrak{h}'$ ist wieder ein H. R. und wir wollen jetzt *allein in ihm operieren.* Seine Schnitte mit $\mathfrak{R}$; $\mathfrak{R}_1, \mathfrak{R}_2, \cdots$ sollen der Reihe nach mit $\mathfrak{R}'$; $\mathfrak{R}_1', \mathfrak{R}_2', \cdots$ bezeichnet werden. Das sind lauter Rotationskörper von $\mathfrak{h}'$ und für ihre Radien gelten die Ungleichungen:

$$\rho_{\mathfrak{R}'} > \sigma_0,$$
$$\rho_{\mathfrak{R}_n'} \leqq \rho_{\mathfrak{R}_n} \leqq \sigma_0.$$

Nun ändern wir unsere Punktmengen $\mathfrak{R}'$; $\mathfrak{R}_1', \mathfrak{R}_2', \cdots$ noch ein wenig ab. $\mathfrak{R}'$ ist in $\mathfrak{h}'$ rotationssymmetrisch in Bezug auf den Schnittpunkt M von $\mathfrak{h}'$ und $\mathfrak{a}_{\mathfrak{R}}$ und enthält den Punkt T mit einer Distanz $\overline{MT} > \sigma_0$. $\mathfrak{K}'$ sei die abgeschlossene Kugel von $\mathfrak{h}'$ mit dem Mittelpunkt M und dem Radius $\overline{MT}$. Wir bilden zunächst die Durchschnitte

$$\mathfrak{R}'' = \mathfrak{K}'\mathfrak{R}'; \qquad \mathfrak{R}_1'' = \mathfrak{K}'\mathfrak{R}_1', \qquad \mathfrak{R}_2'' = \mathfrak{K}'\mathfrak{R}_2', \cdots.$$

Offenbar ist, wieder in $\mathfrak{h}'$,

$$\rho_{\mathfrak{R}''} > \sigma_0 \quad \text{und} \quad \rho_{\mathfrak{R}_n''} \leqq \rho_{\mathfrak{R}_n'} \leqq \sigma_0 \qquad (n = 1, 2, \cdots)$$

und ausserdem

$$\mathfrak{R}'' = \mathfrak{R}_1'' + \mathfrak{R}_2'' + \cdots.$$

Jetzt ersetze man $\mathfrak{R}''$ durch $\mathfrak{K}'$ und ebenso jedes $\mathfrak{R}_n''$ durch den Körper $\mathfrak{S}_n'$ welcher entsteht, indem man zu jedem Punkt von $\mathfrak{R}_n''$ seine ganze Verbindungsstrecke mit M hinzufügt. Wir haben dann

$$\mathfrak{K}' = \mathfrak{S}_1' + \mathfrak{S}_2' + \cdots$$

und offenbar, in $\mathfrak{h}'$, für jedes n

$$\rho_{\mathfrak{S}_n'} = \rho_{\mathfrak{R}_n''},$$

also

$$\rho_{\mathfrak{S}_n'} < \text{Radius von } \mathfrak{K}'.$$

Damit sind wir auf den *Fall der Kugel* zurückgeführt worden. Der Satz ist somit vollständig bewiesen.

Mit Hilfe des eben bewiesenen Hilfssatzes sind wir im Stande, den Begriff des Radius auf allgemeinste beschränkte Punktmengen auszudehnen. Wir

vollziehen die Verallgemeinerung in zwei Etappen. Zunächst gehen wir zu einer Erweiterung des Systems der Rotationskörper über, die später eine wichtige Rolle spielen wird, den *Rotativkörpern. Wir nennen die Punktmenge* $\mathfrak{P}$ *einen Rotativkörper, wenn sie die Darstellung*

$$\mathfrak{P} = \mathfrak{R}_1 + \mathfrak{R}_2 + \cdots \tag{8}$$

zulässt, worin die $\mathfrak{R}_n$ *alle Rotationskörper bedeuten, deren Radien der Bedingung*

$$\lim_{n\to\infty} \rho_{\mathfrak{R}_n} = 0 \tag{9}$$

genügen. Man kann übrigens stets zu einer Darstellung mit *paarweise punktfremden Summanden* übergehen. Zu diesem Zweck setze man

$$\mathfrak{R}_n^* = (\mathfrak{R}_1 + \mathfrak{R}_2 + \cdots + \mathfrak{R}_n) - (\mathfrak{R}_1 + \mathfrak{R}_2 + \cdots + \mathfrak{R}_{n-1}).$$

Dann ist $\mathfrak{R}_n^*$ wieder ein Rotationskörper und

$$\mathfrak{P} = \mathfrak{R}_1^* + \mathfrak{R}_2^* + \cdots, \tag{8*}$$

und ausserdem

$$\mathfrak{R}_i^* \mathfrak{R}_k^* = 0 \qquad (i \neq k). \tag{8**}$$

Da $\mathfrak{R}_n^*$ in $\mathfrak{R}_n$ enthalten ist, ist $\rho_{\mathfrak{R}_n^*} \leqq \rho_{\mathfrak{R}_n}$, also auch

$$\lim_{n\to\infty} \rho_{\mathfrak{R}_n^*} = 0. \tag{9*}$$

Der 8. Hilfssatz legt es nahe, als *Radius* $\rho_{\mathfrak{P}}$ *eines Rotativkörpers* $\mathfrak{P}$ *den Wert*

$$\mu = \text{Max.}\,(\rho_{\mathfrak{R}_1}, \rho_{\mathfrak{R}_2}, \cdots)$$

zu definieren. Allerdings muss bewiesen werden, dass eine andere erlaubte Darstellung von $\mathfrak{P}$ mit gleichen Eigenschaften

$$\mathfrak{P} = \mathfrak{S}_1 + \mathfrak{S}_2 + \cdots$$

zu dem *gleichen Wert* führt.

Beweis: Man setze

$$\nu = \text{Max.}\,(\rho_{\mathfrak{S}_1}, \rho_{\mathfrak{S}_2}, \cdots).$$

Wir führen die Durchschnitte $\mathfrak{T}_{ik} = \mathfrak{R}_i \cdot \mathfrak{S}_k$ ein. Es bestehen die Gleichungen

$$\mathfrak{R}_i = \sum_k \mathfrak{T}_{ik}, \qquad \mathfrak{S}_k = \sum_i \mathfrak{T}_{ik}.$$

Nach dem 8. Hilfssatz ist

$$\rho_{\mathfrak{R}_i} = \text{Max.}\,(\rho_{\mathfrak{T}_{i1}}, \rho_{\mathfrak{T}_{i2}}, \cdots),$$

$$\rho_{\mathfrak{S}_k} = \text{Max.}\,(\rho_{\mathfrak{T}_{1k}}, \rho_{\mathfrak{T}_{2k}}, \cdots).$$

Es ist also

$$\mu = \nu = \text{Maximum aller } \rho_{\mathfrak{T}_{ik}}\,.$$

Für den Bereich der Rotativkörper gelten folgende Sätze:

9'. HILFSSATZ: *Summe und Durchschnitt zweier Rotativkörper ist wieder ein Rotativkörper.*[10]

10'. HILFSSATZ: *Sind* $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_m$ *irgend welche Rotativkörper, so ist*

$$(10') \qquad \rho_{\mathfrak{P}_1+\mathfrak{P}_2+\cdots+\mathfrak{P}_m} = \text{Max.}\,(\rho_{\mathfrak{P}_1}, \rho_{\mathfrak{P}_2}, \cdots, \rho_{\mathfrak{P}_m}).$$

9. HILFSSATZ: *Sind* $\mathfrak{P}_1, \mathfrak{P}_2, \cdots$ *irgend welche Rotativkörper und*

$$\lim_{n\to\infty} \rho_{\mathfrak{P}_n} = 0$$

dann ist auch

$$\mathfrak{P} = \mathfrak{P}_1 + \mathfrak{P}_2 + \cdots$$

ein Rotativkörper.

10. HILFSSATZ: *Unter den Voraussetzungen des 9. Hilfssatzes ist*

$$(10) \qquad \rho_{\mathfrak{P}} = \text{Max.}\,(\rho_{\mathfrak{P}_1}, \rho_{\mathfrak{P}_2}, \cdots).$$

Die Beweise dieser vier Hilfssätze sind so einfach, dass sie dem Leser überlassen werden können.

Nun gehen wir daran, den *Begriff des Radius auf allgemeinste Mengen des H. R. auszudehnen.* Sei $\mathfrak{M}$ eine solche Menge. Gibt es Rotativkörper, welche $\mathfrak{M}$ umschliessen, dann haben deren Radien eine untere Grenze $\rho_{\mathfrak{M}}$. *Wir nennen sie den Radius von* $\mathfrak{M}$. *Wenn es keine umschliessenden Rotativkörper gibt, so schreiben wir* $\mathfrak{M}$ *den Radius* $\rho_{\mathfrak{M}} = \infty$ *zu.*

Mit dieser Definition gelten ganz allgemein die folgenden beiden Hilfssätze:

11'. HILFSSATZ: *Sind* $\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_m$ *irgend welche Mengen, so ist*

$$(11') \qquad \rho_{\mathfrak{M}_1+\mathfrak{M}_2+\cdots+\mathfrak{M}_m} = \text{Max.}\,(\rho_{\mathfrak{M}_1}, \rho_{\mathfrak{M}_2}, \cdots, \rho_{\mathfrak{M}_m}).$$

11. HILFSSATZ: *Sind* $\mathfrak{M}_1, \mathfrak{M}_2, \cdots$ *irgend welche Mengen, welche der Bedingung*

$$\lim_{n\to\infty} \rho_{\mathfrak{M}_n} = 0$$

genügen, so ist

$$(11) \qquad \rho_{\mathfrak{M}_1+\mathfrak{M}_2+\cdots} = \text{Max.}\,(\rho_{\mathfrak{M}_1}, \rho_{\mathfrak{M}_2}, \cdots).$$

Die Beweise ergeben sich fast unmittelbar aus den entsprechenden für Rotativkörper.

BEMERKUNG: Bei der Definition des Radius einer Menge $\mathfrak{M}$ haben wir die Gesamtheit der $\mathfrak{M}$ umschliessenden Rotativkörper herangezogen. Es ist jedoch zu bemerken, dass man sich bei der Bestimmung von $\rho_{\mathfrak{M}}$ auf solche Rotativkörper beschränken kann, welche Vereinigung einer Folge von *Kugeln* mit gegen Null konvergierenden Radien sind. In der Tat lässt sich offenbar ein beschränkter Rotationskörper $\mathfrak{R}$ mit *endlich vielen Kugeln* überdecken, deren Radien

[10] Die Differenzbildung kann aus dem System der Rotativkörper hinausführen.

alle kleiner als eine vorgeschriebene Zahl $\rho' > \rho_{\mathfrak{R}}$ sind. Hieraus folgt, dass man auch einen beliebigen Rotativkörper $\mathfrak{P}$ durch einen aus Kugeln zusammengesetzten umschliessen kann, dessen Radius beliebig wenig von $\rho_{\mathfrak{P}}$ abweicht. Hieraus ergibt sich die Richtigkeit unserer Bemerkung.

3. Das Inhaltsmass bei Rotations-und Rotativkörpern

Bei der Einführung des Inhaltsmasses im H. R. lassen wir uns von dem Cavalierischen Prinzip als heuristischem Prinzip leiten. Wir betrachten zunächst nur Rotationskörper $\mathfrak{R}$, welche den beiden Forderungen genügen:

1. FORDERUNG: Wie bisher soll $\mathfrak{R}$ ein beschränkter Körper sein.

2. FORDERUNG: Ein Meridianschnitt von $\mathfrak{R}$, dessen Dimension um 1 grösser ist als die Dimension k der Achse $\mathfrak{a}_{\mathfrak{R}}$, ergebe eine messbare[11] Punktmenge. Später wird eine *Verschärfung dieser Forderung* eingeführt werden (siehe S. 827).

Wir betrachten *jetzt Meridianschnitte von allen möglichen* Dimensionen. Sie sind offenbar *alle* messbar. Sei $\mathfrak{a}_l$ irgend eine Achse, ihre Dimension sei $l \geqq k$. Schneidet man $\mathfrak{R}$ mit einem zu $\mathfrak{a}_l$ parallelen Raum gleicher Dimension im Abstand $r \geqq 0$,[12] so erhält man eine Punktmenge, welche im Allgemeinen (d. h. abgesehen von einer Menge vor r-Werten vom Masse Null) messbar ist.[13] Dies ist bekanntlich eine Folge der Messbarkeit eines Meridianschnittes von der Dimension $l + 1$ $(l \geqq k)$. Der Inhalt des Schnittes heisse $\mu^l_{\mathfrak{R}}(r)$.

Es ist

$$\mu^l_{\mathfrak{R}}(r) \geqq 0 \quad \text{und } \textit{identisch Null für} \quad r > \rho_{\mathfrak{R}} \tag{12'}$$

und

$$\mu^l_{\mathfrak{R}}(r) \quad \text{integrabel für} \quad r \geqq 0. \tag{12''}$$

Zwischen $\mu^l_{\mathfrak{R}}(r)$-Funktionen, deren Indizes um eins differieren, besteht, wie eine einfache geometrische Überlegung zeigt, die Rekursionsformel

$$\mu^l_{\mathfrak{R}}(r) = 2\int_0^\infty \mu^{l-1}_{\mathfrak{R}}((r^2 + \tau^2)^{\frac{1}{2}})\,d\tau \qquad (l \geqq k + 1). \tag{13'}$$

Führt man die Integrationsvariable $\zeta = (r^2 + \tau^2)^{\frac{1}{2}}$ statt τ ein, so nimmt (13′) die Gestalt

$$\mu^l_{\mathfrak{R}}(r) = 2\int_r^\infty \mu^{l-1}_{\mathfrak{R}}(\zeta)\,\frac{\zeta\,d\zeta}{(\zeta^2 - r^2)^{\frac{1}{2}}}\,{}^{14} \qquad (l \geqq k + 1) \tag{13''}$$

an. Setzt man $\omega^l_{\mathfrak{R}}(x) = \mu^l_{\mathfrak{R}}(x^{\frac{1}{2}})$, so wird (13″) noch einfacher:

$$\omega^l_{\mathfrak{R}}(x) = \int_x^\infty \omega^{l-1}_{\mathfrak{R}}(\xi)\,\frac{d\xi}{(\xi - x)^{\frac{1}{2}}}.{}^{14} \tag{13}$$

---

[11] Messbarkeit von Punktmengen und Integrabilität von Funktionen soll stets im Lebesgueschen Sinne gemeint sein.

[12] Ist $l = 0$, so ist ein Punkt im Abstande r vom Rotationszentrum zu wählen.

[13] Im Fall $l = 0$ ist als Mass eines Punktes der Wert 1 zu nehmen.

[14] Es ist zu beachten, dass als obere Grenze des Integrals (13″) der endliche Wert $\rho_{\mathfrak{R}}$, als obere Grenze des Integrals (13) $\rho^2_{\mathfrak{R}}$ genommen werden kann.

Der Zusammenhang zwischen $\omega_{\mathfrak{R}}^{l}(x)$ und $\omega_{\mathfrak{R}}^{l-1}(x)$ wird also durch eine *Abelsche Integraloperation* vermittelt. Denkt man sich $\omega_{\mathfrak{R}}^{l}(x)$ bekannt, so ist $\omega_{\mathfrak{R}}^{l-1}(x)$ *Lösung einer Abelschen Integralgleichung.* Nun ist mit $\mu_{\mathfrak{R}}^{m}(x)$ auch $\omega_{\mathfrak{R}}^{m}(x)$ für jedes $m \geqq k$ integrabel. Auf der andern Seite ist bekannt, dass die Abelsche Integralgleichung (13) bei gegebener integrabler linker Seite, höchstens eine integrable Lösung haben kann.[15] Hieraus folgt:

1. Satz: *Kennt man in der Reihe $\mu_{\mathfrak{R}}^{k}(r), \mu_{\mathfrak{R}}^{k+1}(r), \cdots$ auch nur ein einziges Glied, so kann man alle übrigen berechnen.*

Wir versuchen jetzt eine naturgemässe Definition der Inhalts*gleichheit* zweier Rotationskörper $\mathfrak{R}$ und $\mathfrak{S}$ zu finden. Wir setzen voraus, dass beide den zwei aufgestellten Forderungen genügen. Wir können also die Funktionenfolge

$$\mu_{\mathfrak{R}}^{k}(r), \mu_{\mathfrak{R}}^{k+1}(r), \cdots \tag{14'}$$

und eine entsprechende für $\mathfrak{S}$ (die Dimension von $\mathfrak{a}_{\mathfrak{S}}$ sei l)

$$\mu_{\mathfrak{S}}^{l}(r), \mu_{\mathfrak{S}}^{l+1}(r), \cdots \tag{14''}$$

bilden. Lässt man sich von dem Cavalierischen Prinzip als heuristischem Prinzip leiten, so kommt man dazu, $\mathfrak{R}$ und $\mathfrak{S}$ *sicherlich dann als inhaltsgleich zu betrachten, wenn für irgendein $m \geqq k, l$*

$$\mu_{\mathfrak{R}}^{m}(r) = \mu_{\mathfrak{S}}^{m}(r) \qquad (r \geqq 0) \tag{15}$$

ist. Wir wissen, dass dann diese Gleichheit für *jedes $m \geqq k, l$* besteht. Spätere Überlegungen (vgl. 6. Satz, S. 830) werden zeigen, dass es zweckmässig ist, *nur dann* die Körper als inhaltsgleich zu betrachten, wenn diese Gleichheit zutrifft. Wir definieren also:

Definition: *$\mathfrak{R}$ und $\mathfrak{S}$ werden inhaltsgleich genannt, wenn für ein $m \geqq k, l$ die Gleichung* (15) *besteht.*

Nun zur Definition des Inhalts*masses*. Die Rotationskörper welche unseren beiden Forderungen genügen, bilden einen *Mengenkörper*; denn sind $\mathfrak{R}$ und $\mathfrak{S}$ zwei Körper des Systems, so ist der Verbindungsraum $\mathfrak{v}$ von $\mathfrak{a}_{\mathfrak{R}}$ und $\mathfrak{a}_{\mathfrak{S}}$ Achse von $\mathfrak{R} + \mathfrak{S}$ und falls $\mathfrak{R} < \mathfrak{S}$ ist, von $\mathfrak{S} - \mathfrak{R}$. Ein Meridian durch $\mathfrak{v}$ liefert also messbare Schnitte sowohl mit $\mathfrak{R}$ als auch mit $\mathfrak{S}$, also auch mit $\mathfrak{S} \pm \mathfrak{R}$. Wenn aber überhaupt ein Meridian eines Rotationskörpers eine messbare Schnittmenge liefert, so auch, wie man sehr leicht einsieht, der kleinstmöglicher Dimension.

Sind $\mathfrak{R}$ und $\mathfrak{S}$ punktfremd, so ergibt die Betrachtung eines gemeinsamen Meridians von der Dimension m

$$\mu_{\mathfrak{R}+\mathfrak{S}}^{m}(r) = \mu_{\mathfrak{R}}^{m} + \mu_{\mathfrak{S}}^{m}(r).$$

Es besteht also *Additivität*. Man könnte deshalb daran denken $\mu_{\mathfrak{R}}^{m}(r)$ für genügend hohes m als Inhaltsmass von $\mathfrak{R}$ einzuführen. Bei Zusammensetzungen

[15] Wie üblich, sind *hier auch wie im Folgenden* zwei Funktionen als identisch anzusehen, wenn sie sich nur in einer Menge von Argumentwerten vom Mass Null unterscheiden.

von Körpern führt aber die Notwendigkeit, den Index m zu variieren, zu Schwierigkeiten. Man kann sie jedoch in folgender Weise umgehen. Man betrachte die Gleichung (13″) und setze darin für l den Wert k ein. Die linke Seite ist die durch den Rotationskörper bestimmte Funktion $\overset{k}{\mu}_{\mathfrak{R}}(r)$. *Wir wollen jetzt die über die 2. Forderung für $\mathfrak{R}$ auf 825 hinausgehende Annahme machen, dass es eine Funktion $\overset{k-1}{\mu}_{\mathfrak{R}}(r)$ gibt, welche in jedem Intervall (ϵ, ∞) mit einer positiven unteren Grenze ϵ integrabel ist und die Abelsche Integralgleichung (13″) löst. Es gibt dann, wie man der Theorie dieser Gleichung entnimmt, genau eine Lösung.* Da $\overset{k}{\mu}_{\mathfrak{R}}(r) = 0$ ist für $r > \rho_{\mathfrak{R}}$, so ist offenbar auch

$$\overset{k-1}{\mu}_{\mathfrak{R}}(r) = 0(r > \rho_{\mathfrak{R}})$$

Nun setze man in (13″) $l = k - 1$. Die linke Seite ist jetzt wieder eine bekannte, in jedem Intervall (ϵ, ∞) integrable Funktion. Wir nehmen wieder an, dass eine in jedem Intervall (ϵ, ∞) *integrable Lösung* $\overset{k-2}{\mu}_{\mathfrak{R}}(r)$ *existiert*. Mit dieser setze man den Prozess fort. Wir nehmen an, dass man so bis zu einer Funktion $\overset{0}{\mu}_{\mathfrak{R}}(r) = \mu_{\mathfrak{R}}(r)$ gelangen kann. Alle Funktionen $\overset{l-1}{\mu}_{\mathfrak{R}}(r)$, $\overset{l-2}{\mu}_{\mathfrak{R}}(r)$, $\cdots$, $\overset{0}{\mu}_{\mathfrak{R}}(r)$ verschwinden für $r > \rho_{\mathfrak{R}}$. *Rotationskörper, welche ausser der Forderung* 1. *auf* S. 825 *diesen zu $\mu_{\mathfrak{R}}(r)$ führenden Prozess gestatten—er stellt eine Verschärfung der 2. Forderung auf S. 825 dar—wollen wir als reguläre Rotationskörper bezeichnen. Wir definieren nun: Die einem regulären Rotationskörper $\mathfrak{R}$ zugeordnete Funktion $\mu_{\mathfrak{R}}(r)$ nennen wir Inhaltsmass oder kurz Inhalt von $\mathfrak{R}$.*

Die Zweckmässigkeit dieser Definition ergibt sich aus den im Folgenden bewiesenen Sätzen.

2. Satz: *Sind $\mathfrak{R}$ und $\mathfrak{S}$ zwei reguläre Rotationskörper mit leerem Durchschnitt,* si ist auch $\mathfrak{R} + \mathfrak{S}$ *regulär*[16] *und es ist*

(16) $$\mu_{\mathfrak{R}+\mathfrak{S}}(r) = \mu_{\mathfrak{R}}(r) + \mu_{\mathfrak{S}}(r).$$

Beweis: Sei $\mathfrak{a}$ eine gemeinsame Achse von $\mathfrak{R}$ und $\mathfrak{S}$ von der Dimension m. Ein Meridian von $\mathfrak{R} + \mathfrak{S}$ durch $\mathfrak{a}$ ist gleichzeitig Meridian von $\mathfrak{R}$, $\mathfrak{S}$ und $\mathfrak{R} + \mathfrak{S}$ und liefert eine messbare Schnittmenge mit $\mathfrak{R} + \mathfrak{S}$. Es ist offenbar

$$\overset{m}{\mu}_{\mathfrak{R}+\mathfrak{S}}(r) = \overset{m}{\mu}_{\mathfrak{R}}(r) + \overset{m}{\mu}_{\mathfrak{S}}(r).$$

Daraus folgt offensichtlich die Regularität von $\mathfrak{R} + \mathfrak{S}$ und die Gleichung

$$\mu_{\mathfrak{R}+\mathfrak{S}}(r) = \mu_{\mathfrak{R}}(r) + \mu_{\mathfrak{S}}(r).$$

3. Satz: *Sind $\mathfrak{R}$ und $\mathfrak{S}$ reguläre Rotationskörper, $\mathfrak{R}$ in $\mathfrak{S}$ enthalten, dann ist auch $\mathfrak{S} - \mathfrak{R}$ regulär und*

(17) $$\mu_{\mathfrak{S}-\mathfrak{R}}(r) = \mu_{\mathfrak{S}}(r) - \mu_{\mathfrak{R}}(r).$$

Der Beweis ist ebenso zu führen wie beim 2. Satze.

Definition: *Unter einem regulären Rotativkörper* $\mathfrak{P}$ verstehe man *einen*

---

[16] Bei nicht leerem Durchschnitt braucht $\mathfrak{R} + \mathfrak{S}$ nicht regulär zu sein.

Körper, welcher sich als Summe

$$\mathfrak{P} = \mathfrak{R}_1 + \mathfrak{R}_2 + \cdots \tag{18}$$

von höchstens abzählbar vielen regulären Rotationskörpern darstelstellen lässt, welche den beiden Bedingungen

$$\mathfrak{R}_i\mathfrak{R}_k = 0 \qquad (i \neq k), \tag{19}$$

$$\lim_{n\to\infty} \rho_{\mathfrak{R}_n} = 0 \tag{20}$$

genügen.

Wir bilden die Reihe

$$\sigma(r) = \mu_{\mathfrak{R}_1}(r) + \mu_{\mathfrak{R}_2}(r) + \cdots \tag{21}$$

Sie konvergiert für jeden positiven Wert von r. In der Tat ist bei vorgegebenem Wert $r_0 > 0$ nur für endlich viele n der Radius $\rho_{\mathfrak{R}_n} > r_0$, nur endlich viele Reihenglieder in (21) sind also für $r > r_0$ von Null verschieden. Es gilt nun der

4. Satz: *Die Summe* $\sigma(r)$ *in* (21) *ändert sich nicht, wenn man von der Darstellung* (18) *zu einer neuen von den gleichen Eigenschaften übergeht.*

Sei etwa

$$\mathfrak{P} = \mathfrak{S}_1 + \mathfrak{S}_2 + \cdots . \tag{18'}$$

Die $\mathfrak{S}_n$ seien alle regulär, ferner

$$\mathfrak{S}_i\mathfrak{S}_k = 0 \qquad (i \neq k), \tag{19'}$$

$$\lim_{n\to\infty} \rho_{\mathfrak{S}_n} = 0. \tag{20'}$$

Man bilde die (21) entsprechende Reihe

$$\tau(r) = \mu_{\mathfrak{S}_1}(r) + \mu_{\mathfrak{S}_2}(r) + \cdots . \tag{21'}$$

Wir führen die Bezeichnung $\mathfrak{T}_{ik} = \mathfrak{R}_i\mathfrak{S}_k$ ein. Alle Durchschnitte $\mathfrak{T}_{ik}$ sind wieder (nicht notwendig reguläre) Rotationskörper und es ist

$$\rho_{\mathfrak{T}_{ik}} \leqq \rho_{\mathfrak{R}_i}, \qquad \rho_{\mathfrak{T}_{ik}} \leqq \rho_{\mathfrak{S}_k}. \tag{22}$$

Sei nun r_0 ein irgendwie vorgegebener positiver Wert. Man bestimme zugehörige Indizes n_1 und n_2 so, dass für jedes

$$n > n_1 \quad \text{der Radius} \quad \rho_{\mathfrak{R}_n} < r_0 \tag{23}$$

und für jedes

$$n > n_2 \quad \text{der Radius} \quad \rho_{\mathfrak{S}_n} < r_0 \tag{23'}$$

ausfällt. Wegen (22) ist auch

$$\rho_{\mathfrak{T}_{ik}} < r_0 \quad \text{für} \quad i > n_1 \quad \text{oder} \quad k > n_2. \tag{24}$$

Man setze ferner

$$\mathfrak{U}_i = \mathfrak{R}_i - \sum_{k=1}^{n_2} \mathfrak{T}_{ik} = \sum_{k=n_2+1}^{\infty} \mathfrak{T}_{ik} \qquad (i = 1, 2, \cdots, n_1) \tag{25}$$

und

$$\mathfrak{B}_k = \mathfrak{S}_k - \sum_{i=1}^{n_1} \mathfrak{T}_{ik} = \sum_{i=n_1+1}^{\infty} \mathfrak{T}_{ik} \qquad (k = 1, 2, \cdots, n_2). \tag{25'}$$

Die $\mathfrak{U}_i$ und $\mathfrak{B}_k$ sind Rotationskörper. Wendet man auf (25) und (25′) den 8. Hilfssatz des vorigen Paragraphen an, so erhält man wegen (24)

$$\rho_{\mathfrak{U}_i} < r_0 \qquad (i = 1, 2, \cdots, n_1), \tag{26}$$

$$\rho_{\mathfrak{B}_k} < r_0 \qquad (k = 1, 2, \cdots, n_2). \tag{26'}$$

Nun betrachten wir wieder die Reihen (21) und (21′), aber nur für $r \geqq r_0$ Aus (23) und (23′) folgt

$$\sigma(r) = \sum_{i=1}^{n_1} \mu_{\mathfrak{R}_i}(r), \qquad \tau(r) = \sum_{k=1}^{n_2} \mu_{\mathfrak{S}_k}(r) \qquad (r \geqq r_0). \tag{27}$$

Sei $\mathfrak{a}$ eine gemeinsame Achse der Rotationskörper

$$\mathfrak{T}_{ik}\ (i = 1, 2, \cdots, n_1 ; k = 1, 2, \cdots, n_2), \qquad \mathfrak{U}_i\ (i = 1, 2, \cdots, n_1),$$
$$\mathfrak{B}_k\ (k = 1, 2, \cdots, n_2).$$

Sie habe die Dimension m. Die Gleichungen (25) und (25′) zeigen zusammen mit den Aussagen (26) und (26′), dass für $r \geqq r_0$ die Gleichungen bestehen

$$\overset{m}{\mu}_{\mathfrak{R}_i}(r) = \sum_{k=1}^{n_2} \overset{m}{\mu}_{\mathfrak{T}_{ik}}(r) \quad \text{für} \quad r \geqq r_0 \qquad (i = 1, 2, \cdots, n_1) \tag{28}$$

und

$$\overset{m}{\mu}_{\mathfrak{S}_k}(r) = \sum_{i=1}^{n_1} \overset{m}{\mu}_{\mathfrak{T}_{ik}}(r) \quad \text{für} \quad r \geqq r_0 \qquad (k = 1, 2, \cdots, n_2), \tag{28'}$$

also

$$\sum_{i=1}^{n_1} \overset{m}{\mu}_{\mathfrak{R}_i}(r) = \sum_{k=1}^{n_2} \overset{m}{\mu}_{\mathfrak{S}_k}(r) \qquad (r \geqq r_0),$$

also auch

$$\sum_{i=1}^{n_1} \mu_{\mathfrak{R}_i}(r) = \sum_{k=1}^{n_2} \mu_{\mathfrak{S}_k}(r) \qquad (r \geqq r_0).$$

Dies zeigt im Verein mit (27) die Identität von $\sigma(r)$ und $\tau(r)$.

Der bewiesene Satz gibt uns die Möglichkeit, den *Inhalt* für einen beliebigen *regulären Rotativkörper* $\mathfrak{P}$ zu definieren: *Wir verstehen darunter die Reihensumme* (21) *für eine Zerlegung von* $\mathfrak{P}$ in *reguläre Rotationskörper* $\mathfrak{R}_n$ $(n = 1, 2, \cdots)$, *welche die Eigenschaften* (19) *und* (20) *besitzen.*

Im Bereich der regulären Rotativkörper gilt der allgemeine Satz:

5. Satz: *Sei* $\mathfrak{P}_1, \mathfrak{P}_2, \cdots$ *eine höchstens abzählbare Reihe von regulären Rotativkörpern und es sei*

$$\mathfrak{P}_i\mathfrak{P}_k = 0 \qquad (i \neq k) \tag{29}$$

und

$$\lim_{n\to\infty} \rho_{\mathfrak{P}_n} = 0. \tag{30}$$

Dann ist auch

$$\mathfrak{P} = \mathfrak{P}_1 + \mathfrak{P}_2 + \cdots$$

ein regulärer Rotativkörper und es ist

$$\mu_{\mathfrak{P}}(r) = \mu_{\mathfrak{P}_1}(r) + \mu_{\mathfrak{P}_2}(r) + \cdots. \tag{31}$$

Der Beweis ist auf Grund des Bisherigen so einfach, dass er dem Leser überlassen bleiben kann.

Wir beweisen nun einen Satz, der eine Rechtfertigung der von uns eingeführten Definition des Inhaltsmasses enthält.

6. Satz: *Wenn zwei reguläre Rotativkörper* $\mathfrak{P}$ *und* $\mathfrak{Q}$ *gleiches Inhaltsmass haben, dann kann man zwei Rotativkörper* $\mathfrak{P}'$ *und* $\mathfrak{Q}'$ *finden derart, dass*

$$\mathfrak{P}' < \mathfrak{P}, \qquad \mathfrak{Q}' < \mathfrak{Q}; \qquad \rho_{\mathfrak{P}-\mathfrak{P}'} = 0, \qquad \rho_{\mathfrak{Q}-\mathfrak{Q}'} = 0$$

ist und dass $\mathfrak{P}'$ *und* $\mathfrak{Q}'$ *eine Zerlegung in paarweise Cavalierisch gleiche Rotationskörper mit gegen Null konvergierenden Radien gestatten.*

Beweis: Seien

$$\mathfrak{P} = \sum_i \mathfrak{R}_i, \qquad \mathfrak{Q} = \sum_i \mathfrak{S}_i$$

Zerlegungen von $\mathfrak{P}$ bzw. $\mathfrak{Q}$ in reguläre Rotativkörper

$$\mathfrak{R}_i\mathfrak{R}_k = 0, \qquad \mathfrak{S}_i\mathfrak{S}_k = 0 \quad (i \neq k); \qquad \lim_{n\to\infty} \rho_{\mathfrak{R}_n} = \lim_{n\to\infty} \rho_{\mathfrak{S}_n} = 0.$$

Nach Voraussetzung ist

$$\mu_{\mathfrak{R}_1}(r) + \mu_{\mathfrak{R}_2}(r) + \cdots = \mu_{\mathfrak{S}_1}(r) + \mu_{\mathfrak{S}_2}(r) + \cdots. \tag{32}$$

Nun wähle man irgendwie eine monoton abnehmende Nullfolge

$$r_1 > r_2 > \cdots; \qquad \lim_{n\to\infty} r_n = 0.$$

Sei $\mathfrak{c}_1$ eine gemeinsame Achse aller $\mathfrak{R}_n$, $\mathfrak{S}_n$, deren Radien mindestens gleich r_1 sind. Ihre Dimension sei m_1. Unter $\mathfrak{R}_n'$ verstehe man den Teil eines solchen $\mathfrak{R}_n$, dessen Punkte von $\mathfrak{c}_1$ mindestens die Distanz r_1 haben. Sonst sei $\mathfrak{R}_n' = 0$. Analog sei $\mathfrak{S}_n'$ zu erklären. Aus (32) folgt

$$\mu_{\mathfrak{R}_1'}^{m_1}(r) + \mu_{\mathfrak{R}_2'}^{m_1}(r) + \cdots = \mu_{\mathfrak{S}_1'}^{m_1}(r) + \mu_{\mathfrak{S}_2'}^{m_1}(r) + \cdots \tag{32'}$$

und hieraus offenbar für die endlich vielen von Null verschiedenen $\mathfrak{R}_m'$, $\mathfrak{S}_n'$

$$\mu^{m_1}_{\mathfrak{R}_1'}(r) + \mu^{m_1}_{\mathfrak{R}_2'}(r) + \cdots = \mu^{m_1}_{\mathfrak{S}_1'}(r) + \mu^{m_1}_{\mathfrak{S}_2'}(r) + \cdots, \tag{33'}$$

oder

$$\mu^{m_1}_{\mathfrak{R}_1'+\mathfrak{R}_2'+\cdots} = \mu^{m_1}_{\mathfrak{S}_1'+\mathfrak{S}_2'+\cdots}. \tag{33}$$

Die Summen $\mathfrak{R}_1^* = \sum \mathfrak{R}_n'$, $\mathfrak{S}_1^* = \sum_n \mathfrak{S}_n'$ bestehen nur aus endlich vielen Gliedern, Sie stellen also Rotationskörper dar, welche nach (33) Cavalierisch gleich sind. Nun verfahre man mit $\mathfrak{P}_1 = \mathfrak{P} - \mathfrak{R}_1^*$, $\mathfrak{Q}_1 = \mathfrak{Q} - \mathfrak{S}_1^*$ ebenso wie bis jetzt mit $\mathfrak{P}$ und $\mathfrak{Q}$, indem man ausserdem r_1 durch r_2 ersetzt. Man erhält so zwei neue Cavalierischgleiche Rotationskörper $\mathfrak{R}_2^*$ und $\mathfrak{S}_2^*$ u.s.w. Man zeigt nun leicht, dass

$$\mathfrak{P}' = \mathfrak{R}_1^* + \mathfrak{R}_2^* + \cdots, \qquad \mathfrak{Q}' = \mathfrak{S}_1^* + \mathfrak{S}_2^* + \cdots$$

Rotativkörper mit Zerlegungen derselben darstellen, wie sie unser Satz fordert.

Zum Schluss dieser Paragraphen stellen wir die Frage, welche Funktionen $\mu(r)$ als *Inhalt von regulären Rotativkörpern* in Betracht kommen und inwiefern sie ein *nichtarchimedisches* Grössensystem bilden.

Drei notwendige Bedingungen für $\mu(r)$ sind aus der Entstehung einer Inhaltsfunktion sofort abzulesen:

(a) *$\mu(r)$ verschwindet ausserhalb eines endlichen Intervalls* $(0, r_0)$ $(r_0 > 0)$

(b) *In jedem Intervall (ϵ, r_0) mit positiver unteren Grenze ϵ ist $\mu(r)$ integrabel.*

(c) *$\mu(r)$ lässt sich in der Form*

$$\mu(r) = \sum_n \mu_n(r)$$

schreiben, wo die $\mu_n(r)$ die Eigenschaften haben:

1) *Die $\mu_n(r)$ erfüllen die Bedingung* (a) *und* (b) *mit Werten r_0^n, welche mit wachsendem n gegen Null konvergieren.*

2) *Eine genügend hohe von n abhängige Iteration der Abelschen Operation verwandelt $\mu_n(r)$ in eine nichtnegative Funktion.*

Wir wollen zeigen, dass die *drei Bedingungen* auch *hinreichend* sind. Die von $\mu_n(r)$ geforderten Eigenschaften zeigen, dass es Inhalt eines regulären Rotationskörper ist, dessen Radius nicht grösser ist als r_0^n. Nun denke man sich eine Folge von parallelen Hyperebenen[17] $\mathfrak{h}_1$, $\mathfrak{h}_2$, $\cdots$ des H. R. im Abstand $d =$ Max. $(r_0^1, r_0^2, \cdots)$. Den zu $\mu_n(r)$ gehörigen Rotationskörper kann man vollständig zwischen die n^{te} und die $(n+1)^{te}$ Hyperebene unterbringen. In dieser Lage sind die Rotationskörper also alle punktfremd, ihre Vereinigung ergibt aber einen Rotativkörper mit dem Inhalt $\mu(r)$.

Die Funktionen, welche die ersten beiden Eigenschaften und von der dritten nur die Teileigenschaft 1 haben, bilden einen *Funktionenmodul*; denn die Prozesse

[17] Unter Hyperbene ist ein vollständiger Orthogonalraum zu einer Richtung zu verstehen.

der Addition und Subtraktion führen aus ihm nicht heraus. Die nicht identisch verschwindenden Funktionen, welche auch noch die vollständige dritte Eigenschaft haben, bilden ein Teilsystem des Moduls, das wohl noch die Addition, aber nicht die Subtraktion gestattet. Man kann sie als die *positiven* Grössen des Moduls bezeichnen. Sind $\mu^1(r)$ und $\mu^2(r)$ zwei Funktionen des Moduls und $\mu^2(r) - \mu^1(r)$ in dem eben angeführten Sinne positiv, so wird man $\mu^1(r)$ *kleiner als* $\mu^2(r)$ bezeichnen. Diese Relation ist transitiv. Es kann aber nicht behauptet werden, dass je zwei Elemente des Moduls vergleichbar sind. Innerhalb des Systems der positiven Grössen gilt offenbar das Archimedische Axiom nicht, *sie bilden ein nichtarchimedisches Zahlsystem.*

4. Inhalte von Körpern, welche sich durch reguläre Rotativkörper approximieren lassen

Sei $\mathfrak{M}$ eine Punktmenge des H. R. von folgenden Eigenschaften: Zu jedem positiven ϵ gebe es einen $\mathfrak{M}$ umschliessenden regulären Rotativkörper derart, dass $\mathfrak{P} - \mathfrak{M}$ einen Radius $\rho_{\mathfrak{P}-\mathfrak{M}} < \epsilon$ hat. Wir zeigen, dass einer solchen Punktmenge in naturgemässer Weise ein Inhalt zugeschrieben werden kann. Sei nämlich $\mathfrak{Q}$ ein zweiter zu demselben ϵ gehöriger regulärer Rotativkörper. Dann unterscheiden sich $\mathfrak{P}$ und $\mathfrak{Q}$ von dem Durchschnitt $\mathfrak{P}\mathfrak{Q}$ um Punktmengen, welche beide in $(\mathfrak{P} - \mathfrak{M}) + (\mathfrak{Q} - \mathfrak{M})$ liegen. Aus $\rho_{\mathfrak{P}-\mathfrak{M}} < \epsilon$ und $\rho_{\mathfrak{Q}-\mathfrak{M}} < \epsilon$ folgt aber

$$\rho_{(\mathfrak{P}-\mathfrak{M})+(\mathfrak{Q}-\mathfrak{M})} < \epsilon. \tag{34}$$

Wegen der Regularität von $\mathfrak{P}$ und $\mathfrak{Q}$ kannn man den Durchschnitt $\mathfrak{P}\mathfrak{Q}$ als Summe von Rotationskörpern schreiben, deren Radien gegen Null konvergieren und *deren Meridianschnitte messbar sind.* Eine genügend hohe Partialsumme $\mathfrak{R}$ erfüllt also die Bedingungen

$$\mathfrak{R} < \mathfrak{P}\mathfrak{Q}, \qquad \rho_{\mathfrak{P}\mathfrak{Q}-\mathfrak{R}} < \epsilon, \tag{35}$$

(36) *Die Meridianschnitte des Rotationskörpers $\mathfrak{R}$ sind messbar.*
Aus (34) und (35) folgt, dass

$$\rho_{\mathfrak{P}-\mathfrak{R}} < \epsilon, \qquad \rho_{\mathfrak{Q}-\mathfrak{R}} < \epsilon \tag{37}$$

ist. Hieraus folgt leicht $\mu_{\mathfrak{P}}^{m}(r) = \mu_{\mathfrak{Q}}^{m}(r)$ für $r > \epsilon$ und genügend hohes m und hieraus auch

$$\mu_{\mathfrak{P}}(r) = \mu_{\mathfrak{Q}}(r) \qquad \text{für } r > \epsilon \tag{38}$$

Wählt man eine Nullfolge von ϵ-Werten und zugehörigen $\mathfrak{P}$, so konvergieren die $\mu_{\mathfrak{P}}(r)$ wie (38) zeigt, gegen eine wohlbestimmte Grenzfunktion $\mu_{\mathfrak{M}}(r)$, die von den Willkürlichkeiten des Prozesses unabhängig ist. *Wir nenen sie den Inhalt von $\mathfrak{M}$ und $\mathfrak{M}$ selbst messbar.*

Mit dieser Definition lässt sich eine Inhaltslehre aufbauen, welche mit der im endlichdimensionalen euklidischen Raume viele Züge gemein hat. Ein

fundamentaler Satz der letzteren ist allerdings nicht erfüllt: Die messbare Mengen bilden nicht einen σ-Körper; denn schon Vereinigung und Durchschnitt von regulären Rotationskörpern braucht nicht regulär zu sein.

In einer späteren Arbeit soll eine Erweiterung der hier gegebenen Inhaltstheorie vorgenommen werden, in welcher die messbaren Mengen einen Körper bilden. Ausserdem soll dort eine Integrationstheorie entwickelt werden.

PRAGUE

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

THE PLATEAU PROBLEM FOR NON-RELATIVE MINIMA

BY MAX SHIFFMAN

(Received March 7, 1939)

1. Introduction

The Problem of Plateau is to find minimal surfaces bounded by a given Jordan curve Γ.[1] Previous investigations[2] have yielded the existence of minimal surfaces bounded by Γ which are either absolute or relative minima in area. The purpose of this investigation is to discuss minimal surfaces which are not relative minima, i.e., of the minimax type.[2a]

The present paper restricts itself to a special class of boundary curves Γ (described in §3), a class which includes curves having a continuously turning tangent line. Under this restriction, we shall prove the following two main results: 1) *if* Γ *bounds two minimal surfaces which are proper relative minima, it bounds at least one minimal surface which is not a proper relative minimum* (*main theorem* I, §4); 2) *the Morse relations apply to the Plateau problem* (*main theorem* II, §10).

The method we shall use for obtaining minimal surfaces, following Douglas, Radó, Courant, is to consider them as extremals for the Dirichlet functional $D[\mathfrak{q}] = \frac{1}{2}\int\int (\mathfrak{q}_u^2 + \mathfrak{q}_v^2)\, du\, dv$ among all surfaces $\mathfrak{q}(u, v)$ bounded by Γ. To prove the first main theorem, let $\mathfrak{q}'$, $\mathfrak{q}''$ be two minimal surfaces bounded by Γ which are proper relative minima. We shall join $\mathfrak{q}'$, $\mathfrak{q}''$ by a connected set C_m of surfaces bounded by Γ on which the least upper bound of $D[\mathfrak{q}]$ is the smallest possible. It is then to be expected that there exists on C_m a minimal surface $\mathfrak{q}$ of the required type. This would be a consequence of the minimizing character of C_m if $D[\mathfrak{q}]$ were a continuous functional; but $D[\mathfrak{q}]$ is merely lower semi-continuous. The major part of this paper consists in overcoming this difficulty.

The methods, especially theorems 4, 5, §§7, 8, developed to prove the first

---

[1] See Radó, "On the Problem of Plateau," Ergebnisse der Math., vol. II, no. 2, 1933, for an excellent account of the Plateau problem and for further literature.

[2] Besides the references in Radó, l.c., see Douglas, "The Problem of Plateau," Bull. Amer. Math. Soc., 1933, pp. 227–251, and "Minimal Surfaces of Higher Topological Structure," Annals of Math., vol. 40, 1939, pp. 205–298; Courant, "Plateau's Problem and Dirichlet's Principle," Annals of Math., vol. 38, 1937, pp. 679–724; Shiffman, "The Problem of Plateau for Minimal Surfaces which are Relative Minima," Annals of Math., vol. 39, 1938, pp. 309–315.

[2a] In the meantime, a paper by Morse and Tompkins on minimal surfaces of general critical type has appeared in the April 1939 issue of this journal. Summaries of their paper and of the present paper are contained in the March and April issues respectively of the Proc. Nat. Acad. Sc., 1939.

main theorem suffice to establish the Morse relations for the Plateau problem. These well-known relations between k-caps and connectivity numbers have been established by Morse in his abstract theory of the variational calculus.[3] It remains to show that on each k-cap there is the desired minimal surface. Theorems 4, 5 serve to prove this.

Sections 2, 3, 4 are of an introductory nature; the considerations in §§5, 6 form the basis of our method; a fundamental deformation theorem is obtained in §7, and the variational condition in §8; finally, the proofs of the main theorems are completed in §§9, 10.

2. The Relevant Spaces

According to the Riemann-Weierstrass theorem,[4] a minimal surface $\mathfrak{q} = \mathfrak{q}(u, v)$ (in vector notation) is characterized by the fact that isometric parameters u, v can be found, i.e., $E = G$, $F = 0$ where $E = \mathfrak{q}_u^2$, $F = \mathfrak{q}_u\mathfrak{q}_v$, $G = \mathfrak{q}_v^2$, and that in terms of these parameters $\mathfrak{q}_{uu} + \mathfrak{q}_{vv} = 0$. The potential character of $\mathfrak{q}(u, v)$ implies that the following function

$$\phi(w) = (\mathfrak{q}_u - i\mathfrak{q}_v)^2 = E - G - 2iF$$

is an analytic function of $w = u + iv$, and the isometric character of u, v that $\phi(w) \equiv 0$. The procedure which we shall adopt for obtaining minimal surfaces is to consider them as the extremal surfaces for the Dirichlet functional[5]

$$D[\mathfrak{q}] = \frac{1}{2}\iint (E + G)\,du\,dv = \frac{1}{2}\iint (\mathfrak{q}_u^2 + \mathfrak{q}_v^2)\,du\,dv = \frac{1}{2}\iint \left(\mathfrak{q}_r^2 + \frac{1}{r^2}\mathfrak{q}_\theta^2\right) r\,dr\,d\theta.$$

Let Γ be a given closed Jordan curve in space. We shall consider surfaces $\mathfrak{q} = \mathfrak{q}(u, v)$ defined over the unit circle of the (u, v)-plane, which map the boundary of the unit circle monotonically on Γ, and which are continuous and have piecewise continuous first derivatives. It will be convenient to use polar coördinates r, θ, and the surfaces will hereafter be expressed as $\mathfrak{q} = \mathfrak{q}(r, \theta)$, $r \leqq 1$. Since the Dirichlet functional is invariant under conformal mapping of the unit circle into itself, we may specify that three given points of the unit circle, θ_1, θ_2, θ_3, be mapped by each surface $\mathfrak{q}(r, \theta)$ into three given points A, B, C of Γ. Surfaces $\mathfrak{q}(r, \theta)$, mapping the boundary $r = 1$ monotonically onto Γ, satisfying this three point condition, and having a finite Dirichlet integral, will be called *admissible* surfaces.

It is convenient to classify the admissible surfaces into various spaces. The distance $|\mathfrak{q}_1 - \mathfrak{q}_2|$ between two surfaces $\mathfrak{q}_1(r, \theta)$, $\mathfrak{q}_2(r, \theta)$ will be chosen as the

---

[3] Morse, "Analysis in the Large," notes of the Institute for Advanced Study, 1936–37, and "Functional Topology and Abstract Variational Theory," Annals of Math., vol. 38, 1937, pp. 386–448.

[4] Cf. Radó, l.c. note 1, chap. 2.

[5] Cf. note 2.

uniform distance,

$$|\mathfrak{q}_1 - \mathfrak{q}_2| = \max_{r \leq 1} |\mathfrak{q}_1(r, \theta) - \mathfrak{q}_2(r, \theta)|.$$

The space of all admissible surfaces with this metric will be denoted by $\mathfrak{S}$. The subspace of all potential surfaces in $\mathfrak{S}$ will be designated by $\mathfrak{P}$. Finally, the space of all surfaces $\mathfrak{q}$ of $\mathfrak{S}$, or of $\mathfrak{P}$, for which

$$D[\mathfrak{q}] \leq N$$

will be designated by $\mathfrak{S}_N$, or $\mathfrak{P}_N$, respectively.

It is clear that $\mathfrak{S}_N \subseteq \mathfrak{S}_{N'}$ if $N < N'$, and $\mathfrak{S} = \sum_N \mathfrak{S}_N = \lim_{N\to\infty} \mathfrak{S}_N$. Similarly for $\mathfrak{P}_N$. We shall now obtain some topological properties of the various spaces introduced. Whereas neither $\mathfrak{S}$ nor $\mathfrak{S}_N$ is compact, we have

LEMMA 1. *The set $\mathfrak{P}_N$ is compact and closed.*

PROOF. Because of the three point condition and of the boundedness of the Dirichlet integral for all $\mathfrak{q}$ in $\mathfrak{P}_N$, the boundary values of all the surfaces in $\mathfrak{P}_N$ are equicontinuous.[6] From any sequence of surfaces in $\mathfrak{P}_N$, therefore, there is a subsequence $\mathfrak{q}_1, \mathfrak{q}_2, \cdots$ whose boundary values converge uniformly. Since $\mathfrak{q}_1, \mathfrak{q}_2, \cdots$ are potential surfaces, they converge uniformly to a potential surface $\mathfrak{q}$, and the derivatives of $\mathfrak{q}_1, \mathfrak{q}_2, \cdots$ converge uniformly in any closed interior domain to the corresponding derivatives of $\mathfrak{q}$. This limit potential surface $\mathfrak{q}$ has boundary values lying monotonically on Γ and satisfies the three point condition. Furthermore, in any circle of radius $\rho < 1$,

$$D_{r \leq \rho}[\mathfrak{q}] = \lim_{n\to\infty} D_{r\leq\rho}[\mathfrak{q}_n] \leq \varliminf_{n\to\infty} D[\mathfrak{q}_n] \leq N;$$

letting $\rho \to 1$, $D[\mathfrak{q}] \leq N$.[7] Hence $\mathfrak{q}$ belongs to $\mathfrak{P}_N$ and the lemma is proved.

The connection between the spaces $\mathfrak{S}$ and $\mathfrak{P}$, $\mathfrak{S}_N$ and $\mathfrak{P}_N$, is given by

LEMMA 2. *There is a deformation of $\mathfrak{S}$ in itself which leaves each element of $\mathfrak{P}$ fixed and deforms $\mathfrak{S}$ into $\mathfrak{P}$. This same deformation deforms the space $\mathfrak{S}_N$ in itself into the space $\mathfrak{P}_N$. Analytically expressed, there is a deformation $f(\mathfrak{q}, t)$ defined and continuous for all $\mathfrak{q}$ in $\mathfrak{S}$ and all t in $0 \leq t \leq 1$ such that*

1) *$f(\mathfrak{q}, t)$ is in $\mathfrak{S}$, $f(\mathfrak{q}, 0) = \mathfrak{q}$, $f(\mathfrak{q}, 1) = \bar{\mathfrak{q}}$ where $\bar{\mathfrak{q}}$ is in $\mathfrak{P}$.*

2) *$f(\mathfrak{q}, t) \equiv \mathfrak{q}$ if $\mathfrak{q}$ is in $\mathfrak{P}$.*

3) *$D[f(\mathfrak{q}, t)]$ is a monotonically decreasing function of t for fixed $\mathfrak{q}$;[8] in particular, $D[\bar{\mathfrak{q}}] \leq D[\mathfrak{q}]$.*

PROOF. Let $\bar{\mathfrak{q}}(r, \theta)$ be the potential surface with the same boundary values as $\mathfrak{q}(r, \theta)$. Define the surface $f(\mathfrak{q}, t)$ for $0 \leq t \leq 1$ by

$$f(\mathfrak{q}, t) = \mathfrak{q} + t(\bar{\mathfrak{q}} - \mathfrak{q}) = \bar{\mathfrak{q}} + (1 - t)\zeta$$

[6] Cf. Courant, l.c. note 2.

[7] This states the lower semi-continuity of $D[\mathfrak{q}]$ in the class of potential surfaces.

[8] Condition 3) is equivalent to the statement that f deforms $\mathfrak{S}_N$ in itself into the space $\mathfrak{P}_N$.

where $\zeta = \mathfrak{q}(r, \theta) - \bar{\mathfrak{q}}(r, \theta)$ and ζ has the boundary values zero. By the minimizing character of potential surfaces with regard to the Dirichlet functional,[9]

$$D[f(\mathfrak{q}, t)] = D[\bar{\mathfrak{q}}] + (1 - t)^2 D[\zeta].$$

Thus, $D[f(\mathfrak{q}, t)]$ is a monotonically decreasing function of t for fixed $\mathfrak{q}$, and $f(\mathfrak{q}, t)$ satisfies properties 1), 2), 3) of the lemma. Furthermore, by a well-known theorem of potential theory, $f(\mathfrak{q}_n, t_n) \rightarrow f(\mathfrak{q}, t)$[10] if $\mathfrak{q}_n \rightarrow \mathfrak{q}$, $t_n \rightarrow t$, so that $f(\mathfrak{q}, t)$ is continuous. The lemma is proved.

Thus, the spaces $\mathfrak{S}$ and $\mathfrak{S}_N$ may be replaced by the essentially equivalent but simpler spaces $\mathfrak{P}$ and $\mathfrak{P}_N$.

3. The Admitted Boundary Curves Γ

In the course of the proofs in §§5, 6, certain restrictions will be made on the boundary curves Γ considered. Let $\mathfrak{q}(\theta)$ be a proper representation of Γ, where $\mathfrak{q}(\theta)$ is a vector function and θ varies over the circumference of the unit circle. We shall require the following two conditions to be satisfied:

1) *$\mathfrak{q}(\theta)$ is of bounded variation.*

2) *There is a δ such that $d\mathfrak{q}(\theta) \cdot d\mathfrak{q}(\phi) \geqq 0$ (a product of vectors!) for all θ, ϕ for which $| \theta - \phi | < \delta$.*[11]

In 2), the relation $d\mathfrak{q}(\theta) \cdot d\mathfrak{q}(\theta) \geqq 0$ shall be shorthand for the statement that there is an ϵ such that

$$[\mathfrak{q}(\theta + \Delta\theta) - \mathfrak{q}(\theta)] \cdot [\mathfrak{q}(\phi + \Delta\phi) - \mathfrak{q}(\phi)] \geqq 0$$

for all positive (or all negative) $\Delta\theta$, $\Delta\phi$ for which $| \Delta\theta | < \epsilon$, $| \Delta\phi | < \epsilon$.

If the two conditions above hold for any one representation of Γ, they hold for all representations of Γ. For, any other representation $\mathfrak{y}(\theta)$ of Γ is obtained from $\mathfrak{q}(\theta)$ by performing a continuous monotonic transformation $\lambda(\theta)$ of the circumference of the unit circle into itself, $\mathfrak{y}(\theta) = \mathfrak{q}(\lambda(\theta))$. Condition 1) is satisfied for $\mathfrak{y}(\theta)$ because of the monotonicity of $\lambda(\theta)$. Condition 2) is satisfied for $\mathfrak{y}(\theta)$ for δ' and ϵ' defined as follows: by virtue of the continuity of $\lambda(\theta)$ there is a δ' such that $| \theta - \phi | < \delta'$ implies $| \lambda(\theta) - \lambda(\phi) | < \delta$, and an ϵ' such that

---

[9] Among all surfaces with given boundary values, the potential surface $\bar{\mathfrak{q}}$ has the smallest Dirichlet integral. If ζ is any surface with boundary values 0 and with a finite Dirichlet integral, and if $\mathfrak{q}_\epsilon = \bar{\mathfrak{q}} + \epsilon\zeta$ then $D[\mathfrak{q}_\epsilon] \geqq D[\bar{\mathfrak{q}}]$. Since

$$D[\mathfrak{q}_\epsilon] = D[\bar{\mathfrak{q}}] + 2\,\epsilon\, D[\bar{\mathfrak{q}}, \zeta] + \epsilon^2 D[\zeta],$$

the usual argument yields $D[\bar{\mathfrak{q}}, \zeta] = 0$, so that

$$D[\mathfrak{q}_\epsilon] = D[\bar{\mathfrak{q}}] + \epsilon^2 D[\zeta].$$

This applies to the case $\zeta = \mathfrak{q} - \bar{\mathfrak{q}}$; for then

$$D[\zeta] \leqq ((D[\mathfrak{q}])^{\frac{1}{2}} + (D[\bar{\mathfrak{q}}])^{\frac{1}{2}})^2 \leqq 4D[\mathfrak{q}].$$

[10] $\rightarrow$ means convergence according to the metric of $\mathfrak{S}$, i.e., uniform convergence.

[11] $| \theta - \phi |$ means the length of the shorter arc joining θ, ϕ.

$|\Delta\theta| < \epsilon'$ implies $|\lambda(\theta + \Delta\theta) - \lambda(\theta)| < \epsilon$. Because of the monotonicity of $\lambda(\theta)$, the quantities $\lambda(\theta + \Delta\theta) - \lambda(\theta)$, $\lambda(\phi + \Delta\phi) - \lambda(\phi)$ have the same sign if $\Delta\theta, \Delta\phi$ have. Thus, for $|\theta - \phi| < \delta'$ and $\Delta\theta, \Delta\phi$ both positive and $< \epsilon'$, we have

$$[\mathfrak{y}(\theta + \Delta\theta) - \mathfrak{y}(\theta)]\cdot[\mathfrak{y}(\phi + \Delta\phi) - \mathfrak{y}(\phi)]$$
$$= [\mathfrak{q}(\lambda(\theta + \Delta\theta)) - \mathfrak{q}(\lambda(\theta))]\cdot[\mathfrak{q}(\lambda(\phi + \Delta\phi)) - \mathfrak{q}(\lambda(\phi))] \geqq 0$$

since $\mathfrak{q}(\theta)$ satisfies 2). Hence 1), 2) hold for $\mathfrak{y}(\theta)$.

Let $\mathfrak{q}(\theta)$ be a given proper representation of Γ satisfying 1), 2). Let $\mathfrak{y}(r, \theta)$ be any surface in $\mathfrak{P}$ with boundary values $\mathfrak{y}(\theta)$, so that $\mathfrak{y}(\theta) = \mathfrak{q}(\lambda(\theta))$. Define the functional $\tau(\mathfrak{y})$ as the smallest number τ such that there are points θ, ϕ for which $|\theta - \phi| = \tau$ and $|\lambda(\theta) - \lambda(\phi)| = \delta$. Then $\mathfrak{y}(\theta)$ satisfies properties 1), 2) with $\tau(\mathfrak{y})$ replacing δ. We prove now that the functional $\tau(\mathfrak{y})$ is lower semi-continuous. Let $\mathfrak{y}_n(r, \theta) \to \mathfrak{y}(r, \theta)$, so that $\lambda_n(\theta) \to \lambda(\theta)$ uniformly; let $\tau(\mathfrak{y}_n)$, at least for a subsequence, converge to t. Then, there are points θ_n, ϕ_n for which $|\theta_n - \phi_n| = \tau(\mathfrak{y}_n)$ and $|\lambda_n(\theta_n) - \lambda_n(\phi_n)| = \delta$. Choosing a subsequence so that $\theta_n \to \theta$, $\phi_n \to \phi$, we get

$$|\theta - \phi| = t \quad \text{and} \quad |\lambda(\theta) - \lambda(\phi)| = \delta.$$

Hence $\tau(\mathfrak{y}) \leqq t$. This proves that $\tau(\mathfrak{y}) \leqq \underline{\lim}\, \tau(\mathfrak{y}_n)$, so that $\tau(\mathfrak{y})$ is lower semi-continuous.

Now, if $\mathfrak{y}$ varies over the closed compact set $\mathfrak{P}_N$, $\tau(\mathfrak{y})$ has a minimum which is attained for some surface $\mathfrak{q}$. Since $\tau(\mathfrak{q})$ must be positive, this proves

LEMMA 3. *There is a positive number τ_N such that any surface $\mathfrak{y}(r, \theta)$ in $\mathfrak{P}_N$ has boundary values $\mathfrak{y}(\theta)$ satisfying:*

$$d\mathfrak{y}(\theta)\cdot d\mathfrak{y}(\phi) \geqq 0 \textit{ for all } \theta, \phi \textit{ for which } |\theta - \phi| < \tau_N .$$

The meaning of conditions 1), 2) for the curve Γ is easily obtained. It is well-known that 1) asserts that Γ has a finite length; 2) states that the angle between the directed secant lines $\mathfrak{q}(\theta_2) - \mathfrak{q}(\theta_1)$ and $\mathfrak{q}(\phi_2) - \mathfrak{q}(\phi_1)$ is at most $\pi/2$ if $|\theta_1 - \phi_1| < \delta$ and if $\theta_2 - \theta_1$ and $\phi_2 - \phi_1$ are both positive and $< \epsilon$. Define the directions of Γ at the point θ as all the limits of the directed secant lines $\mathfrak{q}(\theta_2) - \mathfrak{q}(\theta_1)$ for $\theta_2 > \theta_1$ as $\theta_2 \to \theta$, $\theta_1 \to \theta$. Then 2) above implies that the angles formed by all the directions of Γ at the point θ and all the directions at ϕ are at most $\pi/2$ if θ, ϕ are sufficiently close to each other, $|\theta - \phi| < \delta$. In particular, choosing $\theta = \phi$, all the directions of Γ at a point make angles at most $\pi/2$ with each other. Conversely, if the directions of Γ at a single point make angles less than $\pi/2$ with each other, and if this is true for every point of Γ, then Γ has the property 2).

Curves Γ which have a continuously turning tangent line have the properties 1), 2). These curves are everywhere dense in the space of all Jordan curves.

In what follows, we shall limit ourselves to boundary curves satisfying conditions 1), 2).

4. The First Main Theorem. The Maximum-minimum problem

Our first major result is the following

MAIN THEOREM I. *Let Γ be a Jordan curve satisfying properties 1), 2) of §3. If Γ bounds two minimal surfaces which are proper relative minima, it bounds at least one minimal surface which is not a proper relative minimum.*

The character of the minimum refers to the Dirichlet functional $D(\mathfrak{q})$ in the space $\mathfrak{S}$. The surface $\mathfrak{q}$ is said to be a *proper* relative minimum if $D(\mathfrak{y}) > D[\mathfrak{q}]$ for every surface $\mathfrak{y}$ in $\mathfrak{S}$ different from $\mathfrak{q}$ in a sufficiently small neighborhood of $\mathfrak{q}$.

The proof of the theorem is based on a typical maximum-minimum problem. Let $\mathfrak{q}'$, $\mathfrak{q}''$ be the two minimal surfaces which are proper relative minima. Let C be a closed connected set in $\mathfrak{S}$ containing both $\mathfrak{q}'$ and $\mathfrak{q}''$. Define $d[C]$ to be the least upper bound of $D[\mathfrak{q}]$ for all $\mathfrak{q}$ in C; define d to be the greatest lower bound of $d[C]$ for all such sets C:

$$d[C] = \underset{\mathfrak{q} \text{ in } C}{\text{l.u.b.}}\, D[\mathfrak{q}]$$

$$d = \underset{C}{\text{g.l.b.}}\, d[C].$$

The problem is to find a minimizing closed connected set C_m, i.e., one for which $d[C_m] = d$, and then to establish on C_m the existence of the required minimal surface.

THEOREM 1. *There exists a minimizing closed connected set C_m containing $\mathfrak{q}'$, $\mathfrak{q}''$ and contained completely in $\mathfrak{P}$.*

PROOF. It is necessary to establish the existence of a set C for which $d[C]$ is finite. This will be shown in the course of later work for the special class of boundary curves Γ considered here, and is contained in theorem 3, p. 825.

By lemma 2, any closed connected set C containing $\mathfrak{q}'$, $\mathfrak{q}''$ is deformed into such a set $\bar{C}$ all of whose surfaces belong to $\mathfrak{P}$ and for which $d[\bar{C}] \leqq d[C]$. Therefore the lower bound of $d[C]$ for all C's in $\mathfrak{P}$ is also d. Let $C^1, C^2, \cdots, C^n, \cdots$ be a minimizing sequence of closed connected sets in $\mathfrak{P}$ containing $\mathfrak{q}'$, $\mathfrak{q}''$, $d[C^n] \to d$. Now for sufficiently large N, $d[C^n] \leqq N$ for all n, and all the C^n are contained in $\mathfrak{P}_N$. Construct the set C_m of all limit elements $\mathfrak{q}$ of C^n, i.e. of all $\mathfrak{q}$ such that for some subsequence C^{n_i}, $\mathfrak{q}^{n_i} \to \mathfrak{q}$ where $\mathfrak{q}^{n_i}$ belongs to C^{n_i}. By virtue of the compactness and closedness of $\mathfrak{P}_N$, C_m is a closed connected set in $\mathfrak{P}_N$ containing $\mathfrak{q}'$, $\mathfrak{q}''$. Because of the lower semi-continuity of the Dirichlet functional,

$$D[\mathfrak{q}] \leqq \underline{\lim}\, D[\mathfrak{q}^{n_i}] \leqq \underline{\lim}\, d[C^{n_i}] = d,$$

so that $d[C_m] \leqq d$. But $d[C_m] \geqq d$, and finally $d[C_m] = d$. Theorem 1 is proved.

Note that d is larger than both $D[\mathfrak{q}']$ and $D[\mathfrak{q}'']$. For, C_m contains surfaces $\mathfrak{y}'$ and $\mathfrak{y}''$ arbitrarily close to $\mathfrak{q}'$ and $\mathfrak{q}''$ respectively for which $D[\mathfrak{y}'] > D[\mathfrak{q}']$, $D[\mathfrak{y}''] > D[\mathfrak{q}'']$.

There remains for the proof of the first main theorem to establish the existence

of the required minimal surface on C_m. This would be an immediate consequence of the minimum character of C_m if $D[\mathfrak{q}]$ were a continuous functional. But $D[\mathfrak{q}]$ is only lower semi-continuous. The remainder of this paper is devoted to overcoming this difficulty. We shall prove (theorem 6, §7) that there is a minimal surface $\mathfrak{q}$ on C_m for which $D[\mathfrak{q}] = d$.

5. Expression for the Dirichlet Functional

Let $\mathfrak{q}(r, \theta)$ be a potential surface defined over the unit circle with continuous boundary values $\mathfrak{q}(\theta)$. Let a_n, b_n be the Fourier coefficients of $\mathfrak{q}(\theta)$. A simple calculation shows that the value of the Dirichlet integral of $\mathfrak{q}(r, \theta)$ over a circle of radius ρ, $\rho < 1$, is

$$D_{r\leqq\rho}[\mathfrak{q}] = \frac{\pi}{2}\sum_{n=1}^{\infty} n(a_n^2 + b_n^2)\rho^{2n}.$$

This will be transformed into an expression containing the boundary values $\mathfrak{q}(\theta)$ directly. In the course of the work, the assumptions stated in §3 will be made concerning $\mathfrak{q}(\theta)$. These are finally enumerated at the end of this section in the statement of theorem 2.

Assume that the boundary values $\mathfrak{q}(\theta)$ are continuous and of bounded variation. Then, by integration by parts,

$$a_n = \frac{1}{\pi}\int_0^{2\pi} \mathfrak{q}(\theta)\cos n\theta\, d\theta = -\frac{1}{\pi n}\int_0^{2\pi} \sin n\theta\, d\mathfrak{q}(\theta),$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} \mathfrak{q}(\theta)\sin n\theta\, d\theta = \frac{1}{\pi n}\int_0^{2\pi} \cos n\theta\, d\mathfrak{q}(\theta).$$

Substituting in the expression for $D_{r\leqq\rho}[\mathfrak{q}]$, one obtains

$$D_{r\leqq\rho}[\mathfrak{q}] = \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{\rho^{2n}}{n}\int\int \cos n(\theta - \phi)\, d\mathfrak{q}(\theta)\cdot d\mathfrak{q}(\phi)$$

$$= \frac{1}{2\pi}\int\int\left[\sum_{n=1}^{\infty}\frac{\rho^{2n}}{n}\cos n(\theta - \phi)\right] d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

because of the uniform convergence of $\sum_{n=1}^{\infty}\rho^{2n}/n \cos n(\theta - \phi)$. This series is easily evaluated,

$$\sum_{n=1}^{\infty}\frac{\rho^{2n}}{n}\cos n(\theta - \phi) = \frac{1}{2}\log\frac{1}{(1 - \rho^2)^2 + 4\rho^2\sin^2\frac{1}{2}(\theta - \phi)},$$

by noting that it is the real part of $\sum_{n=1}^{\infty} z^n/n$, where $z = \rho^2 e^{i(\theta - \phi)}$. Therefore

$$D_{r\leqq\rho}[\mathfrak{q}] = \frac{1}{4\pi}\int\int \log\frac{1}{(1 - \rho^2)^2 + 4\rho^2\sin^2\frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$= \frac{1}{4\pi}\int\int \log\frac{1}{\left(\dfrac{1 - \rho^2}{2\rho}\right)^2 + \sin^2\frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

since $\iint d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) = 0$. Performing the limit $\rho \to 1$ under the integral sign, we get[12]

$$D[\mathfrak{q}] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi).$$

To justify this passage to the limit, we shall make the additional assumption that there is a τ such that $d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) \geqq 0$ for all θ, ϕ for which $|\theta - \phi| < \tau$ (see §3). In this, and in what follows, the integration variables θ, ϕ are to vary along the circumference of the unit circle, and $|\theta - \phi|$ is to mean the shorter arc between θ, ϕ.

a) Suppose that

$$\iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

is finite. We have

$$D_{r \leqq \rho}[\mathfrak{q}] = \frac{1}{4\pi}\left(\iint_{|\theta-\phi| \geqq \tau} + \iint_{|\theta-\phi|<\tau} \right) \log \frac{1}{\left(\frac{1-\rho^2}{2\rho}\right)^2 + \sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi).$$

In the first integral, the limit $\rho \to 1$ can be performed under the sign of integration. In the second integral, since (for ρ sufficiently close to 1)

$$0 < \log \frac{1}{\left(\frac{1-\rho^2}{2\rho}\right)^2 + \sin^2 \frac{1}{2}(\theta - \phi)} < \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)}$$

and $d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) \geqq 0$, the limit can also be performed under the sign of integration.[13] Hence

$$D[\mathfrak{q}] = \lim_{\rho \to 1} D_{r \leqq \rho}[\mathfrak{q}] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi).$$

b) Suppose that $D[\mathfrak{q}]$ is finite. Then

$$\begin{aligned}
\iint_{\mu \leqq |\theta-\phi| < \tau} \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)
&= \lim_{\rho \to 1} \iint_{\mu \leqq |\theta-\phi| < \tau} \log \frac{1}{\left(\frac{1-\rho^2}{2\rho}\right)^2 + \sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) \\
&\leqq \lim_{\rho \to 1} \iint_{|\theta-\phi| < \tau} \log \frac{1}{\left(\frac{1-\rho^2}{2\rho}\right)^2 + \sin^2 \frac{1}{2}(\theta - \phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)
\end{aligned}$$

[12] The integral is to be understood as an improper Riemann-Stieltjes integral.
[13] This is a well-known theorem of Lebesgue.

since the integrand is positive for $|\theta - \phi| < \mu < \tau$. Because $D[\mathfrak{q}]$ is finite, this last limit is finite. Letting μ tend to zero, it follows that

$$\iint_{|\theta-\phi|<\tau} \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi)$$

is finite. Hence

$$\frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi)$$

is finite, and case a) shows that it is equal to $D[\mathfrak{q}]$.

This completes the proof of

THEOREM 2. *Let $\mathfrak{q}(r, \theta)$ be a potential surface defined over the unit circle with boundary values $\mathfrak{q}(\theta)$. If*

1) *$\mathfrak{q}(\theta)$ is continuous and of bounded variation,*

2) *there is a τ such that $d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) \geqq 0$ for all θ, ϕ for which $|\theta - \phi| < \tau$, then*

$$D[\mathfrak{q}] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi).$$

6. Lemmas on the Dirichlet Functional

Let $\mathfrak{q}(r, \theta)$, $\mathfrak{y}(r, \theta)$ be two surfaces in $\mathfrak{P}$ with boundary values $\mathfrak{q}(\theta)$, $\mathfrak{y}(\theta)$ respectively. Since the boundary values both lie monotonically on Γ,

$$\mathfrak{y}(\Theta) = \mathfrak{q}(\theta) \quad \text{where} \quad \Theta = \lambda(\theta)$$

and $\lambda(\theta)$ is a monotonic function of θ. Because the boundary values $\mathfrak{q}(\theta)$ or $\mathfrak{y}(\theta)$ may be constant along some arcs, one must be careful concerning the definition of $\lambda(\theta)$. $\lambda(\theta)$ is defined in the following manner. To a given point Q of Γ correspond two arcs, $\theta' \leqq \theta \leqq \theta''$, $\Theta' \leqq \Theta \leqq \Theta''$, which may reduce to points, consisting of all the values of θ, Θ for which $\mathfrak{q}(\theta) = Q$, $\mathfrak{y}(\Theta) = Q$ respectively. If these arcs reduce to points $[\theta' = \theta'', \Theta' = \Theta'']$, set $\lambda(\theta') = \Theta'$; if $\Theta' = \Theta''$, set $\lambda(\theta) = \Theta'$ for all θ in $\theta' \leqq \theta \leqq \theta''$; if $\theta' = \theta''$, $\lambda(\theta)$ has a jump at θ' of amount $\Theta'' - \Theta'$; and finally, set $\lambda(\theta) = \Theta' + \dfrac{\Theta'' - \Theta'}{\theta'' - \theta'}(\theta - \theta')$ if $\theta' \neq \theta''$, $\Theta' \neq \Theta''$. The inverse $\theta = \mu(\Theta)$ to $\Theta = \lambda(\theta)$ is defined in a similar way.

We shall join $\mathfrak{q}(r, \theta)$ and $\mathfrak{y}(r, \theta)$ by a 'linear' family of surfaces all belonging to $\mathfrak{P}$. Define $\mathfrak{y}_\epsilon(r, \theta)$ for ϵ in the interval $0 \leqq \epsilon \leqq 1$ as the potential surface with boundary values $\mathfrak{y}_\epsilon(\theta)$ given by[14] $\mathfrak{y}_\epsilon(\Theta) = \mathfrak{q}(\theta)$ for $\Theta = (1 - \epsilon)\theta + \epsilon\lambda(\theta)$, or, equivalently, by

$$\mathfrak{y}_\epsilon(\Theta) = \mathfrak{y}(\theta) \quad \text{for} \quad \Theta = \epsilon\theta + (1 - \epsilon)\mu(\theta).$$

[14] If $\lambda(\theta)$ has a jump at θ' from Θ' to Θ'', then $\mathfrak{y}_\epsilon(\Theta)$ is defined to be constant, $= \mathfrak{q}(\theta')$, in the interval from $(1 - \epsilon)\theta' + \epsilon\Theta'$ to $(1 - \epsilon)\theta' + \epsilon\Theta''$.

The boundary values of $\mathfrak{y}_\epsilon(r, \theta)$ lie monotonicallly on Γ, the three point condition is satisfied, and $\mathfrak{y}_0(r, \theta) = \mathfrak{q}$, $\mathfrak{y}_1(r, \theta) = \mathfrak{y}$. Now

$$D[\mathfrak{y}_\epsilon] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\Theta - \Phi)}\, d\mathfrak{y}_\epsilon(\Theta)\, d\mathfrak{y}_\epsilon(\Phi).$$

By first setting $\Theta = (1 - \epsilon)\theta + \epsilon\lambda(\theta)$, $\Phi = (1 - \epsilon)\phi + \epsilon\lambda(\phi)$, and then $\Theta = \epsilon\theta + (1 - \epsilon)\mu(\theta)$, $\Phi = \epsilon\phi + (1 - \epsilon)\mu(\phi)$, we obtain

$$D[\mathfrak{y}_\epsilon] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}[(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

and

$$D[\mathfrak{y}_\epsilon] = \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}[(\epsilon(\theta - \phi) + (1 - \epsilon)(\mu(\theta) - \mu(\phi))]}\, d\mathfrak{y}(\theta)\, d\mathfrak{y}(\phi).$$

The first of these forms we shall use for the interval $0 \leqq \epsilon \leqq \frac{1}{2}$, the second for $\frac{1}{2} \leqq \epsilon \leqq 1$.

We shall first prove the following lemmas.

LEMMA 4. *$D[\mathfrak{y}_\epsilon]$ is finite and is a continuous function of ϵ in the closed interval $0 \leqq \epsilon \leqq 1$.*

LEMMA 5. $D[\mathfrak{y}_\epsilon] \leqq N + \dfrac{L^2}{2\pi} \log \dfrac{4}{\sin^2 \frac{1}{2}\tau_N}$, *where N is the larger of the two quantities $D[\mathfrak{q}]$ and $D[\mathfrak{y}]$, τ_N is the quantity defined in lemma 3, and L is the length of the curve Γ.*

LEMMA 6. *$D[\mathfrak{y}_\epsilon]$ has a continuous derivative with respect to ϵ in the open interval $0 < \epsilon < 1$, and*

$$\frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} = -\frac{1}{4\pi} \iint \frac{\lambda(\theta) - \theta - (\lambda(\phi) - \phi)}{\tan \frac{1}{2}[(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi).$$

PROOFS. We have, in the interval $0 \leqq \epsilon \leqq \frac{1}{2}$,

$$D[\mathfrak{y}_\epsilon] = \frac{1}{4\pi}\left(\iint\limits_{|\theta - \phi| \geqq \tau} + \iint\limits_{|\theta - \phi| < \tau} \right) \log \frac{1}{\sin^2 \frac{1}{2}[(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

where $\tau = I_1 + I_2$ is so small that $d\mathfrak{q}(\theta) d\mathfrak{q}(\phi) \geqq 0$ if $|\theta - \phi| < \tau$. Now, for $|\theta - \phi| < \tau < 2\pi/3$ and $0 \leqq \epsilon \leqq \frac{1}{2}$,

$$\tfrac{1}{2} |(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))| \geqq \tfrac{1}{2}(1 - \epsilon)|\theta - \phi| \geqq \tfrac{1}{4}|\theta - \phi|,$$

and this same expression is $\leqq \frac{1}{2}(|\theta - \phi| + \frac{1}{2}\cdot 2\pi) \leqq \pi - \frac{1}{4}|\theta - \phi|$. Therefore

$\sin^2 \frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))] \geqq \sin^2 \frac{1}{4}(\theta-\phi)$, and the integrand in I_2 is

$$\leqq \log \frac{1}{\sin^2 \frac{1}{4}(\theta-\phi)} = \log \frac{1}{\sin^2 \frac{1}{2}(\theta-\phi)} + \log 4 \cos^2 \tfrac{1}{4}(\delta-\phi) \leqq \log \frac{1}{\sin^2 \frac{1}{2}(\theta-\phi)} + \log 4.$$

Since $D[\mathfrak{q}]$ is finite, this is integrable, and I_2 is finite and depends continuously on ϵ.[15] Of course, I_1 is finite and depends continuously on ϵ. Therefore $D[\mathfrak{y}_\epsilon]$ is finite and continuous for $0 \leqq \epsilon \leqq \frac{1}{2}$. Furthermore,

$$I_2 \leqq \frac{1}{4\pi} \iint_{|\theta-\phi|<\tau} \left[\log \frac{1}{\sin^2 \frac{1}{2}(\theta-\phi)} + \log 4\right] d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$\leqq \frac{1}{4\pi} \iint \log \frac{1}{\sin^2 \frac{1}{2}(\theta-\phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$- \frac{1}{4\pi} \iint_{|\theta-\phi|\geqq\tau} \log \frac{1}{\sin^2 \frac{1}{2}(\theta-\phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) + \frac{L^2}{4\pi} \log 4 \leqq N + \frac{L^2}{4\pi} \log \frac{4}{\sin^2 \frac{1}{2}\tau};$$

also,

$$I_1 \leqq \frac{L^2}{4\pi} \log \frac{1}{\sin^2 \frac{1}{4}\tau} \leqq \frac{L^2}{4\pi} \log \frac{4}{\sin^2 \frac{1}{2}\tau}$$

by an easy estimation. Hence, for $0 \leqq \epsilon \leqq \frac{1}{2}$,

$$D[\mathfrak{y}_\epsilon] \leqq N + \frac{L^2}{2\pi} \log \frac{4}{\sin^2 \frac{1}{2}\tau}.$$

Reversing the rôles of $\mathfrak{q}(\theta)$ and $\mathfrak{y}(\theta)$ throughout the whole proof shows that $D[\mathfrak{y}_\epsilon]$ is finite and continuous for $\frac{1}{2} \leqq \epsilon \leqq 1$, and is estimated by the same quantity above. Lemmas 4, 5 are proved.

To prove lemma 6, let ϵ lie in the interval $a \leqq \epsilon \leqq 1-a$, where $a > 0$. Differentiate under the integral sign,

$$\frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} = -\frac{1}{4\pi}\left(\iint_{|\theta-\phi|\geqq\tau} + \iint_{|\theta-\phi|<\tau}\right) \frac{\lambda(\theta)-\theta-(\lambda(\phi)-\phi)}{\tan \frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$= I_3 + I_4.$$

The differentiation is valid for the integral I_3. For I_4, set

$$I_4 = -\frac{1}{4\pi} \iint_{|\theta-\phi|<\tau} \frac{(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))}{\tan \frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]} \cdot \frac{\lambda(\theta)-\theta-(\lambda(\phi)-\phi)}{(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi).$$

---

[15] For, if $\epsilon_n \to \epsilon$ then $D[\mathfrak{y}_{\epsilon_n}] \to D[\mathfrak{y}_\epsilon]$ by the Lebesgue theorem.

The first factor in the integrand is uniformly bounded, and so is the second factor. For, we have

$$\frac{\lambda(\theta) - \theta - (\lambda(\phi) - \phi)}{(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))} = \frac{1}{\epsilon} - \frac{1}{\epsilon} \frac{1}{(1 - \epsilon) + \epsilon \dfrac{\lambda(\theta) - \lambda(\phi)}{\theta - \phi}},$$

so that

$$\frac{1}{a} \geqq \frac{1}{\epsilon} \geqq \frac{\lambda(\theta) - \theta - (\lambda(\phi) - \phi)}{(1 - \epsilon)(\theta - \phi) + \epsilon(\lambda(\theta) - \lambda(\phi))} \geqq \frac{1}{\epsilon} - \frac{1}{\epsilon}\frac{1}{1 - \epsilon} = -\frac{1}{1 - \epsilon} \geqq -\frac{1}{a}.$$

This establishes the validity of the differentiation under the integral sign as well as the continuity of $\dfrac{dD[\mathfrak{y}_\epsilon]}{d\epsilon}$ in $a \leqq \epsilon \leqq 1 - a$.[16] Since $a > 0$ is arbitrary this proves lemma 6.

Thus, $\mathfrak{q}(r, \theta)$ and $\mathfrak{y}(r, \theta)$ have been connected by a path contained entirely in $\mathfrak{P}_N$ for sufficiently large N. This important result was required in the proof of theorem 1, §4, and will be stated in the form of a theorem.

THEOREM 3. *Any two surfaces in $\mathfrak{P}$ can be connected by a continuous path of surfaces contained entirely in $\mathfrak{P}_N$ for sufficiently large N.*

We shall now determine the appearance of the graph of $D[\mathfrak{y}_\epsilon]$ as a function of ϵ for surfaces $\mathfrak{y}$ sufficiently near $\mathfrak{q}$. This requires knowing where maxima and minima occur.

MAIN LEMMA 7. *Let $\mathfrak{q}(r, \theta)$ be any surface in $\mathfrak{P}$ whose boundary values are not constant on any arc,*[17] *and $\mathfrak{y}^{(n)}(r, \theta)$ $n = 1, 2, \cdots$, a sequence of surfaces in $\mathfrak{P}$ tending to $\mathfrak{q}(r, \theta)$. Let $\mathfrak{y}_\epsilon^{(n)}(r, \theta)$, $0 \leqq \epsilon \leqq 1$, be the linear path joining $\mathfrak{q}$ and $\mathfrak{y}^{(n)}$ as in lemmas 4, 5, 6. Then,*

$$\left.\frac{dD[\mathfrak{y}_\epsilon^{(n)}]}{d\epsilon}\right]_{\epsilon=\epsilon^{(n)}} \to 0 \quad as \quad n \to \infty$$

implies that $D[\mathfrak{y}_{\epsilon^{(n)}}^{(n)}] \to D[\mathfrak{y}]$.

PROOF. Let $\mathfrak{q}(\theta)$, $\mathfrak{y}^{(n)}(\theta)$ be the boundary values of $\mathfrak{q}(r, \theta)$, $\mathfrak{y}^{(n)}(r, \theta)$ respectively, so that

$$\mathfrak{y}^n(\Theta) = \mathfrak{q}(\theta) \quad \text{for} \quad \Theta = \lambda^{(n)}(\theta).$$

The condition $\mathfrak{y}^{(n)}(r, \theta) \to \mathfrak{y}(r, \theta)$ implies that $\lambda^{(n)}(\theta)$ converges uniformly to θ.[18] In particular, the jumps of $\lambda^{(n)}(\theta)$ tend to zero as $n \to \infty$. In all the following calculations, the superscript n is omitted.

---

[16] The integrand for $(D[\mathfrak{y}_{\epsilon+\Delta\epsilon}] - D[\mathfrak{y}_\epsilon])/\Delta\epsilon$, $a \leqq \epsilon \leqq 1 - a$, remains bounded as $\Delta\epsilon \to 0$, so that the limit can be taken under the integral sign.

[17] The lemma also holds without this restriction, as follows by a slight modification of the proof.

[18] For, let $\lambda^n(\theta) \to \lambda(\theta)$, at least for a subsequence. From $\mathfrak{y}^{(n)}(\theta) \to \mathfrak{q}(\theta)$ uniformly and $\mathfrak{y}^n(\lambda^n(\theta)) = \mathfrak{q}(\theta)$ we get $\mathfrak{q}(\lambda(\theta)) = \mathfrak{q}(\theta)$. Since $\mathfrak{q}(\theta)$ is not constant on any arc, $\lambda(\theta) \equiv \theta$.

We have

$$D[\mathfrak{y}_\epsilon] - D[\mathfrak{q}] = \frac{1}{4\pi}\left(\iint_{|\theta-\phi|\geqq\tau} + \iint_{|\theta-\phi|<\tau}\right)$$

$$\cdot \log \frac{\sin^2 \frac{1}{2}(\theta-\phi)}{\sin^2 \frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) = I_5 + I_6, \quad \text{and}$$

$$I_6 = \frac{1}{4\pi}\iint_{|\theta-\phi|<\tau} d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)\left\{\log\frac{\sin^2\frac{1}{2}(\theta-\phi)}{(\theta-\phi)^2}\right.$$

$$+ \log \frac{1}{\cos^2 \frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}$$

$$\left. + \log\frac{(\theta-\phi)^2}{\tan^2\frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\right\} = J_1 + J_2 + J_3.$$

The only singular integral is J_3. An estimate of J_3 is, because of $\log x \leqq x$,

$$J_3 \leqq \frac{2}{4\pi}\iint_{|\theta-\phi|<\tau} \frac{\theta-\phi}{\tan\frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) = K.$$

Now,

$$\frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} = -\frac{1}{4\pi}\left(\iint_{|\theta-\phi|\geqq\tau} + \iint_{|\theta-\phi|<\tau}\right)$$

$$\cdot\frac{\lambda(\theta)-\theta-(\lambda(\phi)-\phi)}{\tan\frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi) = I_3 + I_4, \quad \text{and}$$

$$I_4 = -\frac{1}{4\pi\epsilon}\iint_{|\theta-\phi|<\tau}\frac{(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))}{\tan\frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$+\frac{1}{4\pi\epsilon}\iint_{|\theta-\phi|<\tau}\frac{\theta-\phi}{\tan\frac{1}{2}[(1-\epsilon)(\theta-\phi)+\epsilon(\lambda(\theta)-\lambda(\phi))]}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi)$$

$$= \frac{1}{\epsilon}J_4 + \frac{1}{2\epsilon}K.$$

Or, $K = 2\epsilon\dfrac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} - 2\epsilon I_3 - 2J_4$.

Reinsert the superscript n, and let $n \to \infty$, so that $\lambda^{(n)}(\theta)$ converges uniformly to θ. We have:

$$I_3 \to 0,$$

$$J_4 \to -\frac{1}{4\pi}\iint_{|\theta-\phi|<\tau}\frac{\theta-\phi}{\tan\frac{1}{2}(\theta-\phi)}\, d\mathfrak{q}(\theta)\, d\mathfrak{q}(\phi),$$

so that

$$K \to \frac{1}{2\pi} \iint\limits_{|\theta-\phi|<\tau} \frac{\theta - \phi}{\tan \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi);$$

further,

$$I_5 \to 0,$$

$$J_1 = \frac{1}{4\pi} \iint\limits_{|\theta-\phi|<\tau} \log \frac{\sin^2 \frac{1}{2}(\theta - \phi)}{(\theta - \phi)^2} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi),$$

$$J_2 \to \frac{1}{4\pi} \iint\limits_{|\theta-\phi|<\tau} \log \frac{1}{\cos^2 \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi),$$

and

$$\overline{\lim}\, J_3 \leqq \lim K = \frac{1}{2\pi} \iint\limits_{|\theta-\phi|<\tau} \frac{\theta - \phi}{\tan \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi).$$

Combining all these, we obtain

$$\overline{\lim}\, (D[\mathfrak{y}_{\epsilon}^{(n)}{}_{(n)}] - D[\mathfrak{y}]) \leqq \frac{1}{4\pi} \iint\limits_{|\theta-\phi|<\tau} \log \frac{\sin^2 \frac{1}{2}(\theta - \phi)}{(\theta - \phi)^2} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi)$$
$$+ \frac{1}{4\pi} \iint\limits_{|\theta-\phi|<\tau} \log \frac{1}{\cos^2 \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi)$$
$$+ \frac{1}{2\pi} \iint\limits_{|\theta-\phi|<\tau} \frac{\theta - \phi}{\tan \frac{1}{2}(\theta - \phi)} \, d\mathfrak{q}(\theta) \, d\mathfrak{q}(\phi).$$

This inequality being true for all sufficiently small τ, let $\tau \to 0$. Then

$$\overline{\lim}\, (D[\mathfrak{y}_{\epsilon}^{(n)}{}_{(n)}] - D[\mathfrak{q}]) \leqq 0.$$

But

$$\underline{\lim}\, (D[\mathfrak{y}_{\epsilon}^{(n)}{}_{(n)}] - D[\mathfrak{q}]) \geqq 0$$

because of the lower semi-continuity of the Dirichlet functional. Hence,

$$\lim\, (D[\mathfrak{y}_{\epsilon}^{(n)}{}_{(n)}] - D[\mathfrak{q}]) = 0,$$

and lemma 7 is proved.

Lemma 7 will be stated in ϵ, δ form.

LEMMA 8. *Let $\mathfrak{q}(r, \theta)$ be any surface in $\mathfrak{P}$ whose boundary values are not constant on any arc.*[19] *For any η, there is an α and a δ such that, if $\mathfrak{y}(r, \theta)$ is any surface in the δ-neighborhood of $\mathfrak{q}$, then*

$$D[\mathfrak{y}_\epsilon] \geqq D[\mathfrak{q}] + \eta \quad \textit{implies} \quad \left| \frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} \right| \geqq \alpha.$$

[19] The lemma also holds without this restriction. See note 17.

PROOF. If this were not true for some η, there would exist a sequence $\mathfrak{y}^{(n)}$ in $\mathfrak{P} \to \mathfrak{q}$ for which $D[\mathfrak{y}_{\epsilon^{(n)}}^{(n)}] \geqq D[\mathfrak{q}] + \eta$ but $\left.\frac{dD[\mathfrak{y}_\epsilon^{(n)}]}{d\epsilon}\right]_{\epsilon=\epsilon^{(n)}} \to 0$. This contradicts lemma 5.

LEMMA 9. *Let $\mathfrak{q}(r, \theta)$, η, α, δ, $\mathfrak{y}(r, \theta)$ be as in lemma 8. There are two possibilities for the graph of $D[\mathfrak{y}_\epsilon]$ as a function of ϵ in $0 \leqq \epsilon \leqq 1$:*

1) *If $D[\mathfrak{y}] < D[\mathfrak{q}] + \eta$, then $D[\mathfrak{y}_\epsilon]$ is always less than $D[\mathfrak{q}] + \eta$.*

2) *If $D[\mathfrak{y}] \geqq D[\mathfrak{q}] + \eta$, there is an $\bar{\epsilon}$ such that $D[\mathfrak{y}_\epsilon] < D[\mathfrak{q}] + \eta$ for $0 \leqq \epsilon < \bar{\epsilon}$, $D[\mathfrak{y}_{\bar{\epsilon}}] = D[\mathfrak{q}] + \eta$, and $\frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} \geqq \alpha$ for $\bar{\epsilon} \leqq \epsilon \leqq 1$.*

PROOF. By lemma 8, any maxima or minima in the graph of $D[\mathfrak{y}_\epsilon]$ must occur below $D[\mathfrak{q}] + \eta$. Hence 1) follows. Further, if $D[\mathfrak{y}] \geqq D[\mathfrak{q}] + \eta$, there is a first $\bar{\epsilon}$ such that $D[\mathfrak{y}_\epsilon] < D[\mathfrak{q}] + \eta$ for $0 \leqq \epsilon < \bar{\epsilon}$, $D[\mathfrak{y}_{\bar{\epsilon}}] = D[\mathfrak{q}] + \eta$, and $D[\mathfrak{y}_\epsilon] \geqq D[\mathfrak{q}] + \eta$ for $\bar{\epsilon} \leqq \epsilon \leqq 1$. By lemma 8, we must have $\frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon} \geqq \alpha$ in $\bar{\epsilon} \leqq \epsilon \leqq 1$. This completes the proof of lemma 9.

7. The Fundamental Deformation Theorem for the Dirichlet Functional

We shall now construct, on the basis of lemma 8, a deformation which is fundamental for our proof. It is contained in

THEOREM 4. *Let $\mathfrak{q}(r, \theta)$, η, α, δ be as in lemma 8. There is a deformation $f(\mathfrak{y}, t)$ of the space $\mathfrak{P}$ in itself, defined and continuous for all $\mathfrak{y}$ in $\mathfrak{P}$ and $0 \leqq t \leqq 1$, which has the following properties:*

1) *$f(\mathfrak{y}, 0) = \mathfrak{y}$; denote $f(\mathfrak{y}, 1)$ by $\bar{\mathfrak{y}}$.*

2) *$f(\mathfrak{q}, t) \equiv \mathfrak{q}$; $f(\mathfrak{y}, t) \equiv \mathfrak{y}$ if $|\mathfrak{y} - \mathfrak{q}| \geqq \delta$.*

3) *Let $|\mathfrak{y} - \mathfrak{q}| \leqq \delta$. If $D[\mathfrak{y}] \leqq D[\mathfrak{q}] + \eta$, then $D[f(\mathfrak{y}, t)] \leqq D[\mathfrak{q}] + \eta$; if $D[\mathfrak{y}] > D[\mathfrak{q}] + \eta$, then $D[f(\mathfrak{y}, t)]$ is a decreasing function of t for some interval $0 \leqq t \leqq \bar{t}$[20] and thereafter, $\bar{t} \leqq t \leqq 1$, $D[f(\mathfrak{y}, t)] \leqq D[\mathfrak{q}] + \eta$.*

4) Let $|\mathfrak{y} - \mathfrak{q}| = \delta - \sigma\delta$, $0 \leqq \sigma \leqq 1$. *Then either $D[\bar{\mathfrak{y}}] \leqq D[\mathfrak{q}] + \eta$ or $D[\bar{\mathfrak{y}}] \leqq D[\mathfrak{y}] - \sigma\alpha$.*

PROOF. Define $f(\mathfrak{y}, t) \equiv \mathfrak{y}(r, \theta)$ *if* $|\mathfrak{y} - \mathfrak{q}| \geqq \delta$; if $|\mathfrak{y} - \mathfrak{q}| = \delta - \sigma\delta$, $0 \leqq \sigma \leqq 1$, define $f(\mathfrak{y}, t) = \mathfrak{y}_{1-\sigma t}(r, \theta)$, where $\mathfrak{y}_\epsilon(r, \theta)$ is the linear path joining $\mathfrak{q}$ and $\mathfrak{y}$ (and $\epsilon = 1 - \sigma t$). This certainly satisfies 1), 2). It satisfies 3) by virtue of lemma 9. For property 4), suppose that $D[\bar{\mathfrak{y}}] = D[\mathfrak{y}_{1-\sigma}] > D[\mathfrak{q}] + \eta$. By lemma 9, $dD[\mathfrak{y}_\epsilon]/d\epsilon \geqq \alpha$ for $1 - \sigma \leqq \epsilon \leqq 1$, and we have

$$D[\mathfrak{y}] - D[\mathfrak{y}_{1-\sigma}] = \int_{1-\sigma}^{1} \frac{dD[\mathfrak{y}_\epsilon]}{d\epsilon}\, d\epsilon \geqq \alpha\sigma.$$

Thus, 4) is satisfied. Since $f(\mathfrak{y}, t)$ is certainly continuous in $\mathfrak{y}$ and t (because $\mathfrak{y}_\epsilon$ is), the theorem is proved.

[20] $\bar{t}$ depends on $\mathfrak{y}$.

8. The Variational Condition

Theorem 4 is the basic theorem which will eliminate the difficulty that the Dirichlet functional is merely lower semi-continuous and not continuous. In this section, we shall exploit the extremal character of minimal surfaces with regard to the Dirichlet functional. We shall obtain a deformation of the neighborhood of a surface in $\mathfrak{P}$, not a minimal surface, which diminishes the value of the Dirichlet integral.

Let $\mathfrak{q}(r, \theta)$ be a surface in $\mathfrak{P}$ not a minimal surface. The analytic function $\phi(w)$ of $w = u + iv = re^{i\theta}$ defined by

$$\phi(w) = (\mathfrak{q}_u - i\mathfrak{q}_v)^2 = e^{-2i\theta}\left(\mathfrak{q}_r - i\frac{1}{r}\mathfrak{q}_\theta\right)^2$$

cannot vanish identically, for otherwise $\mathfrak{q}(r, \theta)$ would be a minimal surface. The expression for $\phi(w)$ in polar coördinates gives

$$w^2\phi(w) = r^2\left(\mathfrak{q}_r - i\frac{1}{r}\mathfrak{q}_\theta\right)^2 = (r^2\mathfrak{q}_r^2 - \mathfrak{q}_\theta^2) - 2ir\mathfrak{q}_r\mathfrak{q}_\theta,$$

so that

$$-2r\mathfrak{q}_r\mathfrak{q}_\theta = \mathfrak{I}[w^2\phi(w)].$$

Hence the potential function $-2r\mathfrak{q}_r\mathfrak{q}_\theta$ cannot vanish identically.[21] Let (q, γ) be a point in polar coordinates where $-2r\mathfrak{q}_r\mathfrak{q}_\theta \neq 0$. Suppose, for example, that it is negative at this point,

$$-2r\mathfrak{q}_r\mathfrak{q}_\theta = -8b < 0 \quad \text{at} \quad (q, \gamma).$$

Now, if $\mathfrak{y}$ is a surface in $\mathfrak{P}$ close to $\mathfrak{q}$, the derivatives of $\mathfrak{y}$ at (q, γ) are likewise close to the corresponding derivatives of $\mathfrak{q}$ at (q, γ). Choose δ such that for any surface $\mathfrak{y}$ in the δ-neighborhood of $\mathfrak{q}$,

$$-2r\mathfrak{y}_r\mathfrak{y}_\theta \leqq -4b \text{ at } (q, \gamma).$$

Define the surface $\mathfrak{y}_\epsilon(r, \theta)$ by

$$\mathfrak{y}_\epsilon(r, \phi) = \mathfrak{y}(r, \theta) \quad \text{for} \quad \phi = \theta + \epsilon\lambda(r, \theta),$$

where $\lambda(r, \theta)$ is equal, in the annular ring $(1 + q)/2 \leqq r \leqq 1$ to the Poisson kernel

$$\frac{1}{2\pi}\frac{r^2 - q^2}{r^2 - 2rq\cos(\theta - \gamma) + q^2}$$

with pole at (q, γ), and in the circle $r \leqq (1 + q)/2$ to some function which has bounded derivatives and attaches continuously to the values of the Poisson kernel on $r = (1 + q)/2$. Let M be the bounds of the derivatives,

$$|\lambda_r| < M, |\lambda_\theta| < M, \quad \text{and limit } \epsilon \text{ to} \quad |\epsilon| < \frac{1}{2M}.$$

[21] Otherwise $w^2\phi(w) \equiv \text{constant} = 0$ by setting $w = 0$.

Map the unit circle conformally on itself so that the points $\theta_1 + \epsilon\lambda(1, \theta_1)$, $\theta_2 + \epsilon\lambda(1, \theta_2)$, $\theta_3 + \epsilon\lambda(1, \theta_3)$ of the boundary go into θ_1, θ_2, θ_3 respectively, and let $\mathfrak{y}^*_\epsilon(r, \theta)$ be the function resulting out of $\mathfrak{y}_\epsilon(r, \theta)$. Finally, let $\mathfrak{z}_\epsilon(r, \theta)$ be the potential surface with the same boundary values as $\mathfrak{y}^*_\epsilon(r, \theta)$. The boundary values of $\mathfrak{z}_\epsilon(r, \theta)$ lie monotonically on Γ, and the three point condition is satisfied. A simple calculation yields[22]

$$D[\mathfrak{z}_\epsilon] \leqq D[\mathfrak{y}^*_\epsilon] = D[\mathfrak{y}_\epsilon] = \frac{1}{2}\iint \left\{\left(\frac{\partial \mathfrak{y}_\epsilon}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial \mathfrak{y}_\epsilon}{\partial \phi}\right)^2\right\} r\,dr\,d\phi$$

$$= \frac{1}{2}\iint \left\{\left(\mathfrak{y}_r - \frac{\epsilon\lambda_r}{1 + \epsilon\lambda_\theta}\,\mathfrak{y}_\theta\right)^2 + \frac{1}{r^2}\,\frac{\mathfrak{y}_\theta^2}{(1 + \epsilon\lambda_\theta)^2}\right\} r(1 + \epsilon\lambda_\theta)\,dr\,d\theta$$

$$= D[\mathfrak{y}] + \frac{\epsilon}{2}\iint \left\{\lambda_\theta\left(\mathfrak{y}_r^2 - \frac{1}{r^2}\,\mathfrak{y}_\theta^2\right) - 2\lambda_r\,\mathfrak{y}_r\,\mathfrak{y}_\theta\right\} r\,dr\,d\theta$$

$$+ \frac{\epsilon^2}{2}\iint \mathfrak{y}_\theta^2\left(\lambda_r^2 + \frac{1}{r^2}\,\lambda_\theta^2\right)\frac{r\,dr\,d\theta}{1 + \epsilon\lambda_\theta}$$

$$= D[\mathfrak{y}] + \frac{\epsilon}{2}\int_{r=1} \lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta + \epsilon^2 I,$$

where $|\,I\,| < 2M^2D[\mathfrak{y}]$. The single integral is to mean

$$\lim_{\rho\to 1}\int_{r=\rho} \lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta.$$

But, for $\rho > (1 + q)/2$, $\int_{r=\rho} \lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta = (-2r\mathfrak{y}_r\mathfrak{y}_\theta)_{\substack{r=q\\ \theta=\gamma}} \leqq -4b$. The coefficient of ϵ in the expression for $D[\mathfrak{z}_\epsilon]$ is therefore $\leqq -2b$.

Now let $\mathfrak{y}$ belong to $\mathfrak{P}_N$. Choose ϵ positive but $\leqq \tau$, where τ is the smaller of the two quantities $\frac{1}{2M}, \frac{b}{2M^2N}$. Then

$$\frac{1}{2}\int_{r=1} \lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta + \epsilon I \leqq -2b + b = -b.$$

Hence

$$D[\mathfrak{z}_\epsilon] \leqq D[\mathfrak{y}] - b\epsilon \quad \text{for} \quad 0 \leqq \epsilon \leqq \tau.$$

We may define a deformation of $\mathfrak{P}_N$ in itself in the following manner. If $|\,\mathfrak{y} - \mathfrak{q}\,| \geqq \delta$, set $f(\mathfrak{y}, t) \equiv \mathfrak{y}$; if $|\,\mathfrak{y} - \mathfrak{q}\,| = \delta - \sigma\delta$, $0 \leqq \sigma \leqq 1$, set $f(\mathfrak{y}, t) = \mathfrak{z}_{\sigma\tau t}(r, \theta)$. We have $f(\mathfrak{y}, 0) = \mathfrak{z}_0 = \mathfrak{y}$, and $D[f(\mathfrak{y}, t)] = D[\mathfrak{z}_{\sigma\tau t}] \leqq D[\mathfrak{y}] - b\sigma\tau t$. Replacing $b\tau$ by β completes the proof of the following variational condition.

THEOREM 5. *Let $\mathfrak{q}(r, \theta)$ be a surface in $\mathfrak{P}$ not a minimal surface. For any $N > D[\mathfrak{q}]$, there are two positive constants δ and β and a deformation $f(\mathfrak{y}, t)$ of the*

---

[22] See Courant, l.c., note 2.

space $\mathfrak{P}_N$ in itself, defined and continuous for all $\mathfrak{y}$ in $\mathfrak{P}_N$ and $0 \leqq t \leqq 1$, with the following properties:

1) $f(\mathfrak{y}, 0) = \mathfrak{y}$; *denote* $f(\mathfrak{y}, 1)$ *by* $\bar{\mathfrak{y}}$.

2) $f(\mathfrak{y}, t) \equiv \mathfrak{y}$ *if* $|\mathfrak{y} - \mathfrak{q}| \geqq \delta$.

3) *If* $|\mathfrak{y} - \mathfrak{q}| = \delta - \sigma\delta, 0 \leqq \sigma \leqq 1$, *then* $D[f(\mathfrak{y}, t)] \leqq D[\mathfrak{y}] - \beta\sigma t$. *In particular,* $D[\bar{\mathfrak{y}}] \leqq D[\mathfrak{y}] - \beta\sigma$.

9. Proof of the First Main Theorem

We shall now tie together the various threads of our argument. The supposition is that Γ bounds two minimal surfaces $\mathfrak{q}'$, $\mathfrak{q}''$ which are proper relative minima. Furthermore, Γ is restricted to those boundary curves described in §3. By theorem 3, it is possible to connect $\mathfrak{q}'$ and $\mathfrak{q}''$ by a closed connected set C for which the upper bound $d[C]$ of $D[\mathfrak{q}]$ for all $\mathfrak{q}$ on C is finite. By theorem 1, there exists a minimum closed connected set C_m in $\mathfrak{P}$ containing $\mathfrak{q}'$ and $\mathfrak{q}''$, i.e., one for which $d[C_m] = d$, the smallest possible. This minimum value d is larger than both $D[\mathfrak{q}']$ and $D[\mathfrak{q}'']$. We shall now prove, using theorem 4 when $\mathfrak{q}(r, \theta)$ is a minimal surface and theorem 5 when $\mathfrak{q}(r, \theta)$ is not, that on C_m there is a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = d$. To apply theorem 4, the following known lemma is required.

Lemma 10. *A minimal surface $\mathfrak{q}(r, \theta)$ in $\mathfrak{P}$ cannot be constant on any arc of the boundary.*

Proof. Suppose that $\mathfrak{q}(r, \theta) \equiv$ constant $= 0$ on an arc of the unit circle. The minimal surface can then be extended by reflection across this arc by setting $\mathfrak{q}(r, \theta) = -\mathfrak{q}(1/r, \theta)$ if $r > 1$. The arc then lies in the interior of the domain of $\mathfrak{q}(r, \theta)$, so that

$$\mathfrak{q}_r^2 = \frac{1}{r^2}\,\mathfrak{q}_\theta^2 \quad (E = G) \quad \text{on the arc.}$$

Since $\mathfrak{q}_\theta = 0$, it follows that $\mathfrak{q}_r = 0$, and the analytic vector function

$$F(w) = e^{-i\theta}\left(\mathfrak{q}_r - i\frac{1}{r}\,\mathfrak{q}_\theta\right) = 0 \qquad \text{on the arc.}$$

Hence $F(w) \equiv 0$ and $\mathfrak{q}(r, \theta) \equiv$ constant. This contradiction proves the lemma.

Theorem 6. *On the minimum closed connected set C_m containing $\mathfrak{q}'$ and $\mathfrak{q}''$, there exists a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = d$.*

Proof. Suppose that no such minimal surface exists on C_m. Every surface $\mathfrak{q}$ on C_m would then be either a minimal surface for which $D[\mathfrak{q}] < d$, or not a minimal surface at all. Theorems 4, 5 will yield deformations which will diminish the value of $d[C_m]$ in the neighborhood of each surface. By an application of the Heine-Borel-Lebesgue theorem, we shall then obtain a deformation of C_m into C_m' for which $d[C_m'] < d$, contrary to the minimum character of C_m. To apply theorem 5, let N be any number $> d$.

For any $\mathfrak{q}$ on C_m construct in $\mathfrak{P}_N$ the open δ-neighborhood, $N_\delta(\mathfrak{q})$, of $\mathfrak{q}$ where δ is defined as follows: if $\mathfrak{q}$ is not a minimal surface, it is the δ of theorem 5; if $\mathfrak{q}$

Map the unit circle conformally on itself so that the points $\theta_1 + \epsilon\lambda(1, \theta_1)$, $\theta_2 + \epsilon\lambda(1, \theta_2)$, $\theta_3 + \epsilon\lambda(1, \theta_3)$ of the boundary go into θ_1, θ_2, θ_3 respectively, and let $\mathfrak{y}^*_\epsilon(r, \theta)$ be the function resulting out of $\mathfrak{y}_\epsilon(r, \theta)$. Finally, let $\mathfrak{z}_\epsilon(r, \theta)$ be the potential surface with the same boundary values as $\mathfrak{y}^*_\epsilon(r, \theta)$. The boundary values of $\mathfrak{z}_\epsilon(r, \theta)$ lie monotonically on Γ, and the three point condition is satisfied. A simple calculation yields[22]

$$D[\mathfrak{z}_\epsilon] \leqq D[\mathfrak{y}^*_\epsilon] = D[\mathfrak{y}_\epsilon] = \frac{1}{2}\iint\left\{\left(\frac{\partial \mathfrak{y}_\epsilon}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial \mathfrak{y}_\epsilon}{\partial \phi}\right)^2\right\} r\,dr\,d\phi$$

$$= \frac{1}{2}\iint\left\{\left(\mathfrak{y}_r - \frac{\epsilon\lambda_r}{1+\epsilon\lambda_\theta}\mathfrak{y}_\theta\right)^2 + \frac{1}{r^2}\frac{\mathfrak{y}_\theta^2}{(1+\epsilon\lambda_\theta)^2}\right\} r(1+\epsilon\lambda_\theta)\,dr\,d\theta$$

$$= D[\mathfrak{y}] + \frac{\epsilon}{2}\iint\left\{\lambda_\theta\left(\mathfrak{y}_r^2 - \frac{1}{r^2}\mathfrak{y}_\theta^2\right) - 2\lambda_r\mathfrak{y}_r\mathfrak{y}_\theta\right\} r\,dr\,d\theta$$

$$+ \frac{\epsilon^2}{2}\iint \mathfrak{y}_\theta^2\left(\lambda_r^2 + \frac{1}{r^2}\lambda_\theta^2\right)\frac{r\,dr\,d\theta}{1+\epsilon\lambda_\theta}$$

$$= D[\mathfrak{y}] + \frac{\epsilon}{2}\int_{r=1}\lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta + \epsilon^2 I,$$

where $|I| < 2M^2D[\mathfrak{y}]$. The single integral is to mean

$$\lim_{\rho\to 1}\int_{r=\rho}\lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta.$$

But, for $\rho > (1+q)/2$, $\int_{r=\rho}\lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta = (-2r\mathfrak{y}_r\mathfrak{y}_\theta)_{\substack{r=q\\ \theta=\gamma}} \leqq -4b$. The coefficient of ϵ in the expression for $D[\mathfrak{z}_\epsilon]$ is therefore $\leqq -2b$.

Now let $\mathfrak{y}$ belong to $\mathfrak{P}_N$. Choose ϵ positive but $\leqq \tau$, where τ is the smaller of the two quantities $\dfrac{1}{2M}, \dfrac{b}{2M^2N}$. Then

$$\frac{1}{2}\int_{r=1}\lambda(-2r\mathfrak{y}_r\mathfrak{y}_\theta)\,d\theta + \epsilon I \leqq -2b + b = -b.$$

Hence

$$D[\mathfrak{z}_\epsilon] \leqq D[\mathfrak{y}] - b\epsilon \quad \text{for} \quad 0 \leqq \epsilon \leqq \tau.$$

We may define a deformation of $\mathfrak{P}_N$ in itself in the following manner. If $|\mathfrak{y} - \mathfrak{q}| \geqq \delta$, set $f(\mathfrak{y}, t) \equiv \mathfrak{y}$; if $|\mathfrak{y} - \mathfrak{q}| = \delta - \sigma\delta$, $0 \leqq \sigma \leqq 1$, set $f(\mathfrak{y}, t) = \mathfrak{z}_{\sigma\tau t}(r, \theta)$. We have $f(\mathfrak{y}, 0) = \mathfrak{z}_0 = \mathfrak{y}$, and $D[f(\mathfrak{y}, t)] = D[\mathfrak{z}_{\sigma\tau t}] \leqq D[\mathfrak{y}] - b\sigma\tau t$. Replacing $b\tau$ by β completes the proof of the following variational condition.

THEOREM 5. *Let $\mathfrak{q}(r, \theta)$ be a surface in $\mathfrak{P}$ not a minimal surface. For any $N > D[\mathfrak{q}]$, there are two positive constants δ and β and a deformation $f(\mathfrak{y}, t)$ of the*

[22] See Courant, l.c., note 2.

space $\mathfrak{P}_N$ in itself, defined and continuous for all $\mathfrak{y}$ in $\mathfrak{P}_N$ and $0 \leqq t \leqq 1$, with the following properties:

1) $f(\mathfrak{y}, 0) = \mathfrak{y}$; *denote* $f(\mathfrak{y}, 1)$ *by* $\bar{\mathfrak{y}}$.

2) $f(\mathfrak{y}, t) \equiv \mathfrak{y}$ *if* $| \mathfrak{y} - \mathfrak{q} | \geqq \delta$.

3) *If* $| \mathfrak{y} - \mathfrak{q} | = \delta - \sigma\delta, 0 \leqq \sigma \leqq 1$, *then* $D[f(\mathfrak{y}, t)] \leqq D[\mathfrak{y}] - \beta\sigma t$. *In particular,* $D[\bar{\mathfrak{y}}] \leqq D[\mathfrak{y}] - \beta\sigma$.

9. Proof of the First Main Theorem

We shall now tie together the various threads of our argument. The supposition is that Γ bounds two minimal surfaces $\mathfrak{q}'$, $\mathfrak{q}''$ which are proper relative minima. Furthermore, Γ is restricted to those boundary curves described in §3. By theorem 3, it is possible to connect $\mathfrak{q}'$ and $\mathfrak{q}''$ by a closed connected set C for which the upper bound $d[C]$ of $D[\mathfrak{q}]$ for all $\mathfrak{q}$ on C is finite. By theorem 1, there exists a minimum closed connected set C_m in $\mathfrak{P}$ containing $\mathfrak{q}'$ and $\mathfrak{q}''$, i.e., one for which $d[C_m] = d$, the smallest possible. This minimum value d is larger than both $D[\mathfrak{q}']$ and $D[\mathfrak{q}'']$. We shall now prove, using theorem 4 when $\mathfrak{q}(r, \theta)$ is a minimal surface and theorem 5 when $\mathfrak{q}(r, \theta)$ is not, that on C_m there is a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = d$. To apply theorem 4, the following known lemma is required.

Lemma 10. *A minimal surface $\mathfrak{q}(r, \theta)$ in $\mathfrak{P}$ cannot be constant on any arc of the boundary.*

Proof. Suppose that $\mathfrak{q}(r, \theta) \equiv$ constant $= 0$ on an arc of the unit circle. The minimal surface can then be extended by reflection across this arc by setting $\mathfrak{q}(r, \theta) = -\mathfrak{q}(1/r, \theta)$ if $r > 1$. The arc then lies in the interior of the domain of $\mathfrak{q}(r, \theta)$, so that

$$\mathfrak{q}_r^2 = \frac{1}{r^2}\,\mathfrak{q}_\theta^2 \quad (E = G) \quad \text{on the arc.}$$

Since $\mathfrak{q}_\theta = 0$, it follows that $\mathfrak{q}_r = 0$, and the analytic vector function

$$F(w) = e^{-i\theta}\left(\mathfrak{q}_r - i\frac{1}{r}\,\mathfrak{q}_\theta\right) = 0 \qquad \text{on the arc.}$$

Hence $F(w) \equiv 0$ and $\mathfrak{q}(r, \theta) \equiv$ constant. This contradiction proves the lemma.

Theorem 6. *On the minimum closed connected set C_m containing $\mathfrak{q}'$ and $\mathfrak{q}''$, there exists a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = d$.*

Proof. Suppose that no such minimal surface exists on C_m. Every surface $\mathfrak{q}$ on C_m would then be either a minimal surface for which $D[\mathfrak{q}] < d$, or not a minimal surface at all. Theorems 4, 5 will yield deformations which will diminish the value of $d[C_m]$ in the neighborhood of each surface. By an application of the Heine-Borel-Lebesgue theorem, we shall then obtain a deformation of C_m into C_m' for which $d[C_m'] < d$, contrary to the minimum character of C_m. To apply theorem 5, let N be any number $> d$.

For any $\mathfrak{q}$ on C_m construct in $\mathfrak{P}_N$ the open δ-neighborhood, $N_\delta(\mathfrak{q})$, of $\mathfrak{q}$ where δ is defined as follows: if $\mathfrak{q}$ is not a minimal surface, it is the δ of theorem 5; if $\mathfrak{q}$

is a minimal surface for which $D[\mathfrak{q}] < d$, it is the δ of theorem 4 (or lemma 8) for $\eta = \frac{1}{2}(d - D[\mathfrak{q}])$. The property 4) of theorem 4 then reads

$$D[\bar{\mathfrak{y}}] \leqq \frac{D[\mathfrak{q}] + d}{2} \quad \text{or} \quad D[\bar{\mathfrak{y}}] \leqq D[\mathfrak{y}] - \sigma\alpha.$$

These neighborhoods cover the closed compact set C_m, so that a finite number of them, $N_{\delta_\nu}(\mathfrak{q}_\nu)$, $\nu = 1, 2, \cdots, n$, suffice to cover C_m completely. Let T_ν be the deformation of $\mathfrak{P}_N$ given by theorem 4 or theorem 5 corresponding to $\mathfrak{q}_\nu$. Let c be the maximum of $\frac{1}{2}(D[\mathfrak{q}_{\nu'}] + d)$ for all those surfaces $\mathfrak{q}_{\nu'}$ of $\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_n$ which are minimal surfaces, $c < d$. (If there are no minimal surfaces among $\mathfrak{q}_1, \cdots, \mathfrak{q}_n$, then in the following argument either eliminate c or take $c = 0$.) Then each deformation T_ν either does not increase the value of the Dirichlet functional, or does not increase it above c. By the deformation $T_1T_2 \cdots T_{n-1}T_n$,[23] therefore, C_m is mapped into a closed connected set C_{m^*} for which $d[C_{m^*}] \leqq d$.

We assert that $d[C_{m^*}] < d$. If not, there would be a sequence $\mathfrak{y}_1^*, \mathfrak{y}_2^*, \cdots$ belonging to C_{m^*} for which $D[\mathfrak{y}_i^*] \rightarrow d$. These surfaces originate from a sequence $\mathfrak{y}_1, \mathfrak{y}_2, \cdots$ belonging to C_m which converge (by choosing a subsequence) to a surface $\mathfrak{y}$ of C_m. Let $N_{\delta_k}(\mathfrak{q}_k)$ be the first neighborhood which contains $\mathfrak{y}$, and let $| \mathfrak{y} - \mathfrak{q}_k | = \delta_k - 2\sigma\delta_k$ where $0 < 2\sigma \leqq 1$. Let $\mathfrak{y}_1', \mathfrak{y}_2', \cdots$ be the images of $\mathfrak{y}_1, \mathfrak{y}_2, \cdots$ respectively under the deformation $T_1T_2 \cdots T_{k-1}$. Since $\mathfrak{y}$ remains fixed under the deformation, we have $\mathfrak{y}_i' \rightarrow \mathfrak{y}$; hence, for all sufficiently large i, $| \mathfrak{y}_i' - \mathfrak{q}_k | < \delta_k - \sigma\delta_k$. Apply the deformation T_k and let each such $\mathfrak{y}_i'$ be transformed into $\mathfrak{y}_i''$. If $\mathfrak{q}_k$ is a minimal surface, either $D[\mathfrak{y}_i''] \leqq \dfrac{D[\mathfrak{q}_k] + d}{2} \leqq c$, or $D[\mathfrak{y}_i''] \leqq D[\mathfrak{y}_i'] - \alpha_k\sigma \leqq d - \alpha_k\sigma$ by theorem 4; if $\mathfrak{q}_k$ is not a minimal surface, $D[\mathfrak{q}_i''] \leqq D[\mathfrak{y}_i'] - \beta_k\sigma \leqq d - \beta_k\sigma$. In any case, letting c' be the larger of the quantities $c, d - \alpha_k\sigma$ or $c, d - \beta_k\sigma$, we have

$$D[\mathfrak{y}_i''] \leqq c' < d.$$

Finally, the deformation $T_{k+1}T_{k+2} \cdots T_n$ maps $\mathfrak{y}_i''$ into $\mathfrak{y}_i^*$, and $D[\mathfrak{y}_i^*] \leqq c'$. But this contradicts the choice of $\mathfrak{y}_i^*$, for which $D[\mathfrak{y}_i^*] \rightarrow d$. Hence we must have $d[C_{m^*}] < \text{d}$.

Under the deformation $T_1T_2 \cdots T_n$ the minimal surfaces $\mathfrak{q}', \mathfrak{q}''$ describe paths on which $D[\mathfrak{q}] \leqq c'' < d$, where c'' is the largest of the three quantities $D[\mathfrak{q}']$, $D[\mathfrak{q}'']$, c. Adding these paths to C_{m^*} yields a closed connected set $C_{m'}$ containing $\mathfrak{q}', \mathfrak{q}''$ for which $d[C_{m'}] < d$. But we must have $d[C_{m'}] \geqq d$. This is a contradiction, and our original supposition is false. Theorem 6 is proved.

The minimal surface $\mathfrak{z}$ on C_m for which $D[\mathfrak{z}] = d$ cannot be a proper relative minimum, since there are surfaces of C_m in any arbitrary neighborhood of $\mathfrak{z}$, and for each surface $\mathfrak{y}$ of C_m $D[\mathfrak{y}] \leqq d = D[\mathfrak{z}]$. The proof of main theorem 1, stated in §4, is complete.

---

[23] Applied in the order from left to right.

10. The Morse Relations

The methods, especially theorems 4, 5, which have been developed here to prove the first main theorem also serve to establish the Morse relations.[24] In the terminology of Morse, a k-cap in the space $\mathfrak{P}$ with cap limit a is a k-dimensional Vietoris chain[25] on $\mathfrak{P}_a$ which is a cycle d-mod $\mathfrak{P}_a$ (this is read definitely mod $\mathfrak{P}_a$ and means mod $\mathfrak{P}_{a'}$ for some $a' < a$), and which is not homologous to zero on $\mathfrak{P}_a$ d-mod $\mathfrak{P}_a$. In particular, the k-cap can not be deformed on $\mathfrak{P}_a$ to a point set d-on $\mathfrak{P}_a$. It follows from theorems 4, 5,[26] exactly as in the proof of theorem 6, that on each k-cap with cap limit a there is a minimal surface $\mathfrak{q}$ for which $D[\mathfrak{q}] = a$.

Let M_k be a maximal group of k-caps with rank.[27] If each M_k, $k = 0, 1, \cdots$, has a finite dimension μ_k, then the Morse relations state

$$\mu_0 \geqq R_0$$

$$\mu_1 - \mu_0 \geqq R_1 - R_0$$

$$\vdots$$

$$\mu_n - \mu_{n-1} + \cdots + (-1)^n \mu_0 \geqq R_n - R_{n-1} + R_{n-2} - \cdots + (-1)^n R_0$$

$$\vdots$$

where $R_0, R_1, \cdots, R_n, \cdots$ are the connectivity numbers of $\mathfrak{P}$. These connectivity numbers are understood in the sense that only chains which lie on $\mathfrak{P}_N$ for some sufficiently large N (depending on the chain) are admitted.

Decompose the set of all minimal surfaces $\mathfrak{q}$ in $\mathfrak{P}$ for which $D[\mathfrak{q}] =$ constant $= a$ into its maximal connected components (as subsets of $\mathfrak{P}$), and call each such component a *bloc* of minimal surfaces. Associated to any bloc σ of minimal surfaces are k-caps with cap limit a which lie in an arbitrary neighborhood of σ.[28] The dimension of a maximal group of k-caps with rank associated with σ is called by Morse the k^{th} type number of σ. Because each k-cap with cap limit a contains a minimal surface $\mathfrak{q}$ for which $D[\mathfrak{q}] = a$, it follows that μ_k is the sum of the k^{th} type numbers of all blocs of minimal surfaces.

It remains to compute the connectivity numbers of $\mathfrak{P}$.

THEOREM 7. *Let $R_0, R_1, \cdots, R_n, \cdots$ be the connectivity numbers of* $\mathfrak{P}$. *Then* $R_0 = 1$, $R_1 = R_2 = \cdots = R_n = \cdots = 0$.

---

[24] References to the extensive literature on the Morse relations will be found in Morse, "The Calculus of Variations in the Large," Amer. Math. Soc. Coll. Publ., vol. 18 and "Functional Topology etc.," l.c., note 3. The procedure and terminology which we shall follow are contained in the references cited in note 3.

[25] Vietoris, "Über den höheren Zusammenhang kompakten Räume und eine Klasse von zusammenhangenstreuen Abbildungen," Math. Ann. vol. 97, 1927, pp. 454–472. See Morse, "Functional Topology etc.," l.c., note 3.

[26] Theorem 4 is a more general property than the concept of "reducibility" as introduced by Morse.

[27] Cf. Morse, l.c., note 3.

[28] Cf. Morse, l.c., notes 3 and 24.

PROOF. The first relation $R_0 = 1$ is the content of theorem 3. Let z^n, $n \geqq 1$, be any n-dimensional singular cycle in $\mathfrak{P}_N$. Choose any surface $\mathfrak{q}$ in $\mathfrak{P}_N$, and connect each surface $\mathfrak{y}$ in z^n to $\mathfrak{q}$ by the linear path $\mathfrak{y}_t$ of lemmas 4, 5. This defines a singular $(n + 1)$-chain whose boundary is z^n and which lies on $\mathfrak{P}_{N'}$ for

$$N' = N + \frac{L^2}{2\pi} \log \frac{4}{\sin^2 \frac{1}{2}\tau_N},$$

by lemma 5. Hence, z^n is homologous to zero, and $R_n = 0$. It is clear what modifications to make in this proof if Vietoris chains are used in place of singular chains.

We have therefore established, on the basis of the Morse theory, the

MAIN THEOREM II. *Let μ_k be the sum of the k^{th} type numbers of all blocs of minimal surfaces bounded by Γ (with the restrictions on Γ stated in §3). Under the assumption that each μ_k, $k = 0, 1, \cdots$, is finite, the following Morse relations hold*:

$$\begin{aligned} \mu_0 &\geqq 1 \\ \mu_1 - \mu_0 &\geqq -1 \\ &\vdots \\ \mu_n - \mu_{n-1} + \cdots + (-1)^n\mu_0 &\geqq (-1)^n \\ &\vdots \end{aligned}$$

Several problems remain in connection with this theorem. First, to show that a bloc of minimal surfaces bounded by Γ consists of one minimal surface. Second, to show that the k'th type number of a minimal surface bounded by Γ is either zero for all k, or zero for all $k \neq j$ and 1 for $k = j$. Third, to characterize the Morse type of a minimal surface by properties similar to the number of conjugate points (or the number of negative characteristic roots in the associated problem) in the case of single integral problems in the calculus of variations.[29]

NEW YORK UNIVERSITY.

[29] Cf. Morse, l.c., notes 3 and 24, where similar questions are considered for single integral problems. Also, with reference to minimal surfaces, cf. the classical work of Schwarz, collected works, vol. I, pp. 151–167, 223–269.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

A THEOREM CONCERNING ANALYTIC CONTINUATION FOR FUNCTIONS OF SEVERAL COMPLEX VARIABLES

BY A. E. TAYLOR[1]

(Received March 6, 1939)

1. The Main Theorem

The purpose of this note is to prove the following remarkable theorem pertaining to analytic functions of several complex variables.

THEOREM. *Let T be a bounded open set with connected boundary C, in the $2n$-dimensional (real) Euclidean space of the n complex variables $x_1, \cdots, x_n$ $(n > 1)$. Let $f(x_1, \cdots, x_n) \equiv f(x)$ be a single-valued function which is analytic [meromorphic] in some region containing C. Then it is possible to extend f, by analytic continuation, to a function which is single-valued and analytic [meromorphic] throughout $T + C$ (that is, in some open set containing $T + C$).*

The bracketed words afford an alternative reading of the theorem as a statement about meromorphic functions.[2]

The theorem, when f is analytic, is due to Hartogs; for meromorphic functions it was enunciated by E. E. Levi.[3] The following proof grew out of a study of the demonstrations of the theorem given by Osgood and by A. B. Brown.[4] The method is materially different from that of either of these men, however. It is the belief of the author that the proof to be given, rather long though it is in detail, is fundamentally extremely simple and easy to grasp intuitively. The detail seems necessary to avoid unjustifiably hasty conclusions.

2. The First Fundamental Lemma

The theorem to be proved stems from the following lemma.

LEMMA 1. *Let P be a boundary point of a sphere K in the $2n$-dimensional space under consideration $(n > 1)$, and let f be a function which is analytic [meromorphic] in that portion of a neighborhood of P which lies outside K. Then it is possible to continue f across the boundary of K in the neighborhood of P so that the extended function is analytic [meromorphic] at P. (Here again the bracketed words afford an alternative reading).*[5]

---

[1] Part of the work on this paper was done while the author was a National Research Fellow at Princeton University.

[2] For the definition of a meromorphic function of several variables see Osgood, *Lehrbuch der Funktionen theorie*, II, 1 (1929), pp. 180–183.

[3] The theorems, with references to the literature, are to be found in Osgood, loc. cit. p. 206 and p. 220.

[4] The paper by A. B. Brown is in the Duke Mathematical journal, vol. 2 (1936) p. 20 ff.

[5] For the proof of this lemma, with the two readings, see Osgood, loc. cit. p. 204 and p. 215.

Lemma 1 and the main theorem are false when $n = 1$, so that they are distinctly propositions about functions of *several* variables.

The only uses which we shall make of the theory of analytic functions in the proof will be the appeal to Lemma 1 and to the fact that an analytic [meromorphic] function is uniquely determined by its values in an arbitrary open set within its domain of definition. The rest of the argument is concerned with the nature of certain point sets. We shall first outline the general plan of the proof, and then proceed with the details. We shall deal throughout with the case that f is analytic. No modification of the proof is needed for the alternative theorem, when f is meromorphic.

Under the hypothesis f is analytic in a spherical neighborhood about each point of C, and we can select a finite number of these open spheres such that their point set sum S contains C. Clearly S is a bounded, connected, open set, and so is $T + S$. We introduced the notations $T_0 = T + C, E = T + S$. We shall prove that f may be extended analytically to all of E.

Choose a point P_0 not in the closure of E and consider spheres about this point as center. There will obviously exist such a sphere (call its bounding surface Σ_0) with points of E outside Σ_0, points of C on Σ_0, and no point of T outside Σ_0. For as the radius r_0 of Σ_0 we can take the maximum distance between P_0 and a variable point of C. The function f is the analytic in the portion of E outside Σ_0. Now it is evident that the sphere Σ_0 could be made a little smaller and f would continue to be defined and analytic in the part of E outside the sphere. If this shrinking process were continued, however, we should ultimately have points of E on the sphere at which f would not be defined. For any individual obstruction of this kind Lemma 1 provides us with a remedy: we can extend f across the sphere at this point provided that f is defined at all the nearby points outside.

Two questions now arise: (1) at any stage in the process of shrinking the sphere will it always be possible to shrink it more, and still have an analytic continuation of f defined at all the points of E outside the new sphere? (2) May there not be some sphere to which we can approach arbitrarily near with our slightly larger spheres, but which itself can never be attained in the shrinking process?

By setting forth a definite procedure for passing from one sphere to a smaller one, all the while maintaining an analytic continuation of f into an open set containing all the points of E outside the sphere, we shall deal with the difficulties suggested by the above questions, and show that they may be overcome.

3. The Second Fundamental Lemma

First we shall define the concept of an *attainable sphere.*

Definition: *A sphere with surface Σ and center P_0 will be called attainable under the following circumstances. There exists a nonempty open set G outside Σ, together with a function φ single-valued and analytic in G. G and φ shall have the further properties*: (1) *G contains all points of E outside Σ*; (2) *a boundary*

point of G outside Σ shall belong to the set B of boundary points of S not in E;
(3) *The points of the boundary of G on Σ fall into several classes, as described below. We distinguish two of these classes by the letters D, R, and a part of D by F.*
(a) *D includes $\Sigma \cdot T_0$;*
(b) *R is composed of all points of $\Sigma \cdot (E - T_0)$ not in D;*
(c) *D is closed, $D + R$ is open relative*[6] *to Σ;*
(d) *the set F of points of D which are limit points of R is a subset of C;*
(e) *All the points near a point of $D + R$ and outside Σ are in G;*
(f) *the boundary points of G on Σ but not in $D + R$ are in B.*
(4) *$\varphi = f$ at all the points near a point of R and outside Σ.*

It may be easily verified that the sphere Σ_0 defined in §2 is attainable if we take G_0 to be the part of E outside Σ_0, $\varphi_0 = f$ in G_0, $D_0 = \Sigma_0 \cdot T_0$, and $R_0 = \Sigma_0(E - T_0)$.

Lemma 2. *If Σ is attainable and D is non-empty there exists a smaller attainable sphere Σ', the corresponding set G' and function φ' standing in the relations $G' \supset G$, $\varphi' = \varphi$ in G.*

Proof: By Lemma 1 and (3e) there exists, corresponding to each point P of D, a small sphere with center at P such that φ admits an analytic continuation into the interior of the sphere. Since D is closed and bounded we may take a uniform radius $\eta > 0$ for all such spheres. If r is the radius of Σ and δ the minimum distance between C and the boundary of S we may choose ϵ positive, less than $\eta/2$, δ and r, and define Σ' as the sphere with center at P_0 and radius $r' = r - \epsilon$. Then

I. *All points of E between Σ, Σ' and a distance $\geqq \epsilon$ from D are in S.* For the radius from P_0 through such a point P could not meet Σ in a point of D (since $r - r' = \epsilon$). If the point were in $E - S$ (and so in T) the radius would either have to meet Σ in a point of T_0 (which is impossible, by (3a)), or cross C between P and Σ. In the latter case P would be a distance $< \epsilon < \delta$ from C, and hence in S.

We now define G'. It shall consist of all points of G, D, R, together with all points between Σ and Σ' which are either (α) a distance $< \epsilon$ from D, or (β) points of E a distance $\geqq \epsilon$ from D (such points are in S, by *I*). We leave it to the reader to verify that G' is open, non-empty, and that it contains all points of E outside Σ'.

II. *If P is a point on or inside Σ then either* (i) *the minimum distance from P to points of D is attained at a point of F, or* (ii) *P is inside Σ and the radius from P_0 through P meets Σ in a point of D such that all the nearby points on Σ are also in D.* This is a quite general proposition to which we shall frequently refer. It depends rather obviously on the geometry of the sphere, together with property (3c) and the definition of F. We leave detailed verification to the reader.

[6] That is, $D + R$ is such that if P is in it all the points on Σ and sufficiently near to P are also in $D + R$.

III. *A boundary point of G' outside Σ' and on or inside Σ is a distance $\geqq \delta$ from D, and is therefore in B.* For if P is such a point its minimum distance from D is attained at a point of F (otherwise it would be in G', by II and the definition of G'). Since F is a subset of C, by (3d), P would be in S if its distance from F were less than δ. This is impossible. Now if P were inside Σ it would be the limit of points of G' a distance $> \epsilon$ from D (since $\epsilon < \delta$) and hence in S, by I. Thus P would be in B. If P were on Σ it would be the limit of points of G' outside, on, or inside Σ. In the latter two cases it would be in B by an argument similar to the one just given. In the first case it would be in B by (3f).

Next we define φ'. For points of G we define $\varphi' = \varphi$. For points of R and points of S between Σ, Σ' and a distance $\geqq \epsilon$ from D we put $\varphi' = f$. For a point of D or a point between Σ, Σ', a distance $< \epsilon$ from D we define φ' by the analytic continuation of φ into the neighborhood of points of D, as already mentioned. Since $\epsilon < \eta/2$ the definition of φ' at the points of this last class is unique, by the uniqueness of analytic continuation.

In verifying that φ' is analytic in G' the only points at which special attention is required are (i) points of R a distance $< \epsilon$ from D, and (ii) points of R a distance ϵ from D, and points between Σ, Σ', a distance ϵ from D. For points P of class (i) we observe that R is open relative to Σ and that $\varphi' = f$ at all points near P on or outside Σ (by (4)), while φ' is an analytic continuation of φ at the points near P inside Σ. Hence $\varphi' = f$ in the full neighborhood of P. For a point P of class (ii) the nearest point Q of D is in F (and so in C) by II, (3d). Since $\overline{PQ} = \epsilon < \delta$ it follows that the entire neighborhood of P is in S, as is also the interior of the sphere of radius ϵ about Q. Then $\varphi' = f$ in this sphere, since there are points of R near Q (by definition of F). Thus $\varphi' = f$ in the neighborhood of P. This completes the proof that φ' is analytic in G'.

To complete the proof of the lemma we must define D', R', and show that Σ' has the properties (3), (4) of an attainable sphere.

DEFINITION: *R' is the set of points of $\Sigma' \cdot (E - T_0)$ for which $\varphi' = f$ at the nearby points outside Σ'. D' is the set of boundary points of G' on Σ' which are either a distance ϵ from D, but not in R', or are in $\Sigma' \cdot T_0$.*

A boundary point of G' on Σ' not in $D' + R'$ must be a distance $> \epsilon$ from D, and hence in B, by *I*. It follows from this that R' contains all points of $\Sigma' \cdot (E - T_0)$ not in D'. Thus we are left to verify properties 3c, d, e. To verify that $D' + R'$ is open relative to Σ' it suffices to consider a point P of D', since clearly R' is open relative to Σ'. If P is an $\Sigma' \cdot T_0$ it is clear that all the nearby points on Σ' are in E, hence in $D' + R'$. If P is a distance ϵ from D the line P_0P meets Σ in a point Q of D. If Q is in F, P is in S (since $F \subset C$ and $\epsilon < \delta$). If Q is in $D - F$ the points near Q on Σ are in D and hence the points near P on Σ' are in D'. Thus $D' + R'$ is open relative to Σ'. This argument also shows that all the points near a point of $D' + R'$ and outside Σ' are in G'.

D' is closed. For let $\{P_n\}$ be a sequence of points in D', converging to a

limit P. If an infinite number of the P_n are in $\Sigma' \cdot T_0$ so is P, since T_0 is closed. If an infinite number of the P_n are a distance ϵ from D, say $\overline{P_n Q_n} = \epsilon$, Q_n in D, we may assume that the Q_n have a limit Q (in D, since it is closed). But then we easily see that $\overline{PQ} = \epsilon$ and hence P is in D', for it cannot be in R', which is open relative to Σ'.

Finally, F', the set of limit points of R' in D', is a subset of C. For let $P_n \to P$, P in D', P_n in R'. If P is in $\Sigma' \cdot T_0$ it is necessarily in C because of the nature of R'. If this is not the case P is a distance ϵ from D, and the nearest point of D is in F (other-wise the points near P on Σ' would all be in D', as we saw when we proved that $D' + R'$ is open relative to Σ'). Thus P is in S. It cannot be in $E - T_0$ without being in R' because of property (4) and the fact that $P_n \to P$, P_n in R'. Hence P must be in T_0, and on C. This completes the verification of the fact that Σ' is an attainable sphere, and finishes the proof of the lemma.

4. Proof of the Main Theorem

We have seen that there exists an attainable sphere Σ_0 with a non-empty set D_0. The process of Lemma 2 then leads us to a smaller attainable sphere Σ_1. If Σ_1 intersects C the set D_1 is non-empty, and we can obtain a still smaller attainable sphere, and so on. Now it is clear that the theorem will be proved if we can prove the existence of an attainable sphere which is smaller than Σ_0, and which does not intersect C (the connectedness of C enters here).

The repetition of the process of Lemma 2 is capable of providing us with such an attainable sphere. For if it were not there would exist a sequence of attainable spheres $\Sigma_0, \Sigma_1, \cdots$ with radii $r_0 > r_1 > \cdots$, all the Σ_i intersecting C, and the r_i having a limit $r > 0$ such that the sphere Σ of radius r and center P_0 would be non-attainable. Furthermore, each Σ_{i+1} and its associated G_{i+1}, φ_{i+1}, etc. would be related to those of its predecessor Σ_i in the manner set forth in the proof of Lemma 2.

We shall prove that under the circumstances just described the sphere Σ *is* attainable, thus yielding a contradiction. We begin by defining $G = G_0 + G_1 + \cdots$, $\varphi = \varphi_i$ in G_i. Evidently φ is single-valued and analytic in G. We leave it for the reader to verify that Σ enjoys the properties (1), (2) of an attainable sphere (as defined in §3). It is then clear that a boundary point of G on Σ which is such that in every neighborhood of it there are points outside Σ *and not in* G is a limit point of B; hence it is itself in B. We shall call such points points of class I.

The remaining boundary points of G on Σ, which we shall call points of class II, have the property that the nearby points outside Σ are all in G.

Definition. *R is the totality of points of class* II *which are in $E - T_0$ and such that $\varphi = f$ at the nearby points outside Σ. D is the remaining part of class* II, *except for limit points of R which are in B.*

The only properties of D, R not immediately obvious are the properties 3c, d. We turn our attention to these.

Since R is open relative to Σ we may, in proving that $D + R$ is open relative to Σ, consider points of D only. If P is such a point all the nearby points on Σ are in class II. They are then either in $D + R$, or are limit points of R in B. But points of the latter kind form a closed set, and if there were some of them in every neighborhood of P, P itself would be a point of this kind, contrary to its being in D. Thus $D + R$ is open relative to Σ.

F (as defined in (3d)) is a subset of C. For let P be in F. Being a limit point of R, P is either in $E - T_0$, on C, or a limit point of R in B. The latter cannot be the case, since P is in D. If it were in $E - T_0$ it would be in R because being a limit point of R we could infer that $\varphi = f$ at the points near P outside Σ. Hence P must be on C.

Finally, D is closed. For let P be a limit point of D. P cannot be in R, since R is open relative to Σ. Hence if P is not in D it must, by (3f), be in B, for the boundary of G is a closed set. Two cases arise, according as P is in class I or II.

P in Class I: In this case P is the limit of a sequence of points of B lying on the boundary of G outside Σ. Now S is composed of the sum of a finite number of open spheres. It follows from this that B is such that there is a whole arc of boundary points of G, outside Σ with P as an end point. In particular, it follows that we may choose points Q_i along this arc, Q_i between Σ_i and Σ_{i+1}, for *all* sufficiently large values of i, and such that $Q_i \to P$. Then Q_i is a distance $\geqq \delta$ from D_i, and $\varphi_{i+1} = f$ at the points of G_{i+1} near Q_i. This follows from II, §3, the fact that Q_i is in B, and the way in which φ_{i+1}, G_{i+1} are defined. We shall show that all the points of G in a sufficiently small neighborhood of P are in $E - T_0$, and that $\varphi = f$ at these points. To do this choose j so that $r_j - r < \delta/4$, and so that Q_i, $i \geqq j$, exists, with $\overline{Q_iP} < \delta/4$. Let K be the interior of the sphere of radius $r_j - r$ and center at P. Observe that $r_i - r_{i+1} = \epsilon_i < \delta/4$ if $i \geqq j$. For a fixed $i \geqq j$ a point of $K \cdot G_{i+1}$ on or inside Σ_i is a distance $\leqq \delta/2$ from Q_i, which is a distance $\geqq \delta$ from D_i. Hence the point in question is a distance $\geqq \delta/2 > \epsilon_i$ from D_i. It is then in S, and $\varphi_{i+1} = f$ there. This argument shows that all the points of G in K are in S, and that $\varphi = f$ there. Since P is in B all these points must be in $E - T_0$.

We now recall that P is a limit point of D, so that we have $P_n \to P$, P_n in D. When n is sufficiently large P_n will be in K. Thus all the points near P_n outside Σ, being in G, since P_n is in class II, are in $E - T_0$, and $\varphi = f$ there. It then follows that the points near P on Σ are either in $E - T_0$ or in B (they cannot be on C because of proximity to B). Now not all of them can be on B, because of the nature of B, which is composed of parts of the boundaries of the spheres which make up S. Hence P_n is a point B which is a limit point of R. This contradicts the fact that it is in D.

P in Class II: The only remaining possibility is that P is in class II. In this case it is a limit point of R as well as of D. Let K be a small neighborhood of P such that all the points of K outside Σ are in G. Then the intersection of K with a Σ_i must consist entirely of points of $D_i + R_i$. If K is sufficiently

small there can be no points of F_i in the intersection, for P is a distance $\geqq \delta$ from C. By taking K smaller, if necessary, we can insure that the intersection with Σ_i contains points of R_i, for in any neighborhood of P there is a point of R, and $\varphi = f$ at the nearly points of G. But if a portion of the intersection is in R_i it all is, as otherwise there would be a point of F_i in it.

Now not *all* the points in K can be in $E - T_0$, for if $P_n \to P$, P_n in D, this would lead, as we have just seen a moment ago, to the conclusion that some P_n is not in D. There must therefore be a point Q of B in K and outside Σ. We know that Q is not on any Σ_i. Suppose it is between Σ_i and Σ_{i+1}, and so in G_{i+1}. By I, §3 and the definition of G_{i+1} its distance from D_i is $\leqq \epsilon_i = r_i - r_{i+1}$. The nearest point of D_i cannot be in F_i, for we can assume K very small, and this would bring P too near C. Thus, by II, §3 the radius of Σ_i passing through Q meets Σ_i in D_i, contrary to the fact that Σ_i intersects K solely in R_i. Thus we are again led to a contradiction. P is then necessarily in D, D is closed, Σ is an attainable sphere, and the proof of the theorem is complete.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

AN INITIAL VALUE PROBLEM FOR ALL HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS OF SECOND ORDER WITH THREE INDEPENDENT VARIABLES[1]

BY EDWIN W. TITT

(Received December 30, 1938)

This paper deals with an initial value problem that can be set for any analytic hyperbolic equation (rank 2 or 3) linear or non-linear in three independent variables with any inclination of the initial surface. This initial value problem is a Cauchy problem in which we require that the initial data be analytic in one of their arguments but not in the other. Special instances of this problem have been considered by Volterra[2] and Hadamard on the one hand and Hamel[3] on the other.

The $(n - 1)$ dimensional characteristic strips employed in this paper are an extension on the one hand of a concept well known in the theory of the second order equation in two independent variables[4] and on the other hand of a concept which occurs in the theory of the first order equation in more than two independent variables and which has not been clearly recognized.[5]

The method used in handling our initial value problem in a great many places bears a resemblance to the method used by Lewy[6] in showing that Cauchy's problem for a non-linear hyperbolic equation in two independent variables has a unique solution. In the proof of our existence theorems we

---

[1] Developed in part while the author was a National Research Fellow and presented to the American Mathematical Society, Dec. 31, 1935. See Abstract 42-1-40, *Bull. Amer. Math. Soc.*, vol. 42 (1936), p. 32. The present treatment incorporates many improvements and additions. The author wishes to take this opportunity to express his appreciation to Professor G. A. Bliss for his interest and encouragement when this work was in its early stages.

[2] See J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale Univ. Press, New Haven, (1923), p. 254.

[3] G. Hamel, *Dissertation*, Göttingen, 1901.

[4] See Goursat, *Leçons sur l'intégration des équations aux dérivées partielles du second ordre*, Tome I, Hermann, Paris, (1896), pp. 171–4.

[5] See E. W. Titt, "$(n - 1)$—Dimensional Characteristic Strips of a First Order Equation and Cauchy's Problem," *Duke Math. Jour.*, vol. 3 (1937), pp. 740–6. *Cf.* Courant u. Hilbert, *Methoden der Mathematischen Physik*, II, Julius Springer, Berlin, (1937), pp. 82–7.

[6] H. Lewy, "Über das Anfangswertproblem bei einer hyperbolischen nichtlinearen partiellen Differentialgleichung zweiter Ordnung mit zwei unabhängigen Veränderlichen," *Math. Annalen*, vol. 98 (1927–28), pp. 179–191. An exposition of Lewy's work has been given by J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*, Hermann, Paris, (1932), pp. 487–508. See also *loc. cit.*, Courant u. Hilbert, Kap. V.

employ a combination of the method of successive approximations and the method of dominant functions. Although Lewy used a method of difference equations in his original paper, his existence theorems have since been handled by the method of successive approximations.[7] Our use of dominant functions in the existence proof for linear equations resembles somewhat Hadamard's[8] use of dominant functions in the proof of the convergence of the elementary solution.

For a detailed formulation of our problem and the plan of the paper the reader is referred to the first part of §2.

Before proceeding to the discussion of our problem we wish to state the definition of a partially analytic function and mention certain of its properties. A function $f(x, y)$ of two real variables x and y will be said to be partially analytic with respect to y for $y = y_0$ in the interval $\alpha \leqq x \leqq \beta$ provided it can be represented by a series of the form

$$f(x, y) = a_0(x) + a_1(x)(y - y_0) + a_2(x)(y - y_0)^2 + \cdots, \tag{i}$$

whose coefficients are continuous functions of x in the interval $\alpha \leqq x \leqq \beta$ and provided that the series (i) converges absolutely and uniformly for $\alpha \leqq x \leqq \beta$, $|y - y_0| \leqq \gamma$. Sometimes in what follows it will be convenient to refer to a set of points satisfying inequalities of the type $\alpha \leqq x \leqq \beta$, $|y - y_0| \leqq \gamma$ as the region of partial analyticity.

In particular we have that $f(x, y)$ is a continuous function of x and y for $\alpha \leqq x \leqq \beta$, $|y - y_0| \leqq \gamma$ and the integral $\int_\alpha^x f(x'y)\, dx'$ can be evaluated using term by term integration and the integrated series will again converge absolutely and uniformly for $\alpha \leqq x \leqq \beta$, $|y - y_0| \leqq \gamma$. Due to the absolute and uniform convergence we are able to obtain a function,

$$\frac{m}{1 - \dfrac{y - y_0}{r}},$$

in which m and r are positive constants, which dominates the series (i) for all x which satisfy $\alpha \leqq x \leqq \beta$. Due to the presence of this dominant function the series obtained from (i) using term by term differentiation with respect to y will again converge absolutely and uniformly in any subregion $\alpha \leqq x \leqq \beta$, $|y - y_0| \leqq \delta < \gamma$. Suppose that we have another function $g(x, z)$ par-

[7] In this connection the author is deeply indebted to H. Lewy. In 1935 the author had the privilege of reading Lewy's lectures on partial differential equations delivered at Göttingen in 1927–28 (unpublished). Substantially the same treatment of the existence theorems is given in Courant u. Hilbert, *loc. cit.*, pp. 317–326.

[8] See *loc. cit.*, Hadamard (1923), Chap. III. For an exposition see T. Y. Thomas and E. Titt, "On the Elementary Solution of the General Linear Differential Equation of the Second Order with Analytic Coefficients," to appear in the *Jour. de Math. Pures et App.*

tially analytic with respect to z in a region $\alpha \leqq x \leqq \beta$, $|z - z_0| \leqq \epsilon$. If $|g(x, z_0) - y_0| < \gamma$, then the result of replacing y in $f(x, y)$ by g will be another function $F(x, z)$ partially analytic with respect to z in some region $\alpha \leqq x \leqq \beta$, $|z - z_0| \leqq \eta < \epsilon$. The extension to more variables is evident and other properties such as implicit function theorems will cause no difficulties.

1. Cauchy's problem in general

Consider the partial differential equation

$$F(p_{\alpha\beta} \mid p_\alpha \mid z \mid x^\alpha) = 0 \qquad (\alpha,\beta = 1, 2, 3) \tag{1.1}$$

in one unknown z and three independent variables x^α, where the p_α and $p_{\alpha\beta}$ denote the first and second partial derivatives respectively of z with respect to x^α. Cauchy's problem for the equation (1.1) is the problem of passing a solution through an *initial strip*

$$\begin{aligned} x^\alpha &= \xi^\alpha(u, w); & z &= \zeta(u, w); \\ p_\alpha &= \pi_\alpha(u, w); & p_{\alpha\beta} &= \pi_{\alpha\beta}(u, w); \end{aligned} \tag{1.2}$$

satisfying (1.1) and subject to the strip conditions

$$\begin{aligned} z_u &= p_\alpha x^\alpha_u\,; & z_w &= p_\alpha x^\alpha_w\,; \\ \frac{\partial p_\alpha}{\partial u} &= p_{\alpha\beta} x^\beta_u\,; & \frac{\partial p_\alpha}{\partial w} &= p_{\alpha\beta} x^\beta_w\,; \end{aligned} \tag{1.3}$$

where the subscripts u and w denote partial differentiation with respect to u and w.

Let us inquire into the conditions under which the equation (1.1) will determine the third derivatives $p_{\alpha\beta\gamma}$ over the initial strip. Differentiating the equation (1.1) with respect to x^γ considering all of its arguments as functions of x^α, we obtain

$$a^{\alpha\beta} p_{\alpha\beta\gamma} + b_\gamma = 0, \tag{1.4}$$

where

$$a^{\alpha\alpha} = \frac{\partial F}{\partial p_{\alpha\alpha}}, \qquad 2a^{\alpha\beta} = \frac{\partial F}{\partial p_{\alpha\beta}} \qquad (\alpha \neq \beta),$$

$$b_\gamma = \frac{\partial F}{\partial p_\alpha}\, p_{\alpha\gamma} + \frac{\partial F}{\partial z}\, p_\gamma + \frac{\partial F}{\partial x^\gamma}.$$

In addition to (1.4) the derivatives $p_{\alpha\beta\gamma}$ must satisfy the conditions

$$\frac{\partial p_{\alpha\gamma}}{\partial u} = p_{\alpha\beta\gamma} x^\beta_u\,; \qquad \frac{\partial p_{\alpha\gamma}}{\partial w} = p_{\alpha\beta\gamma} x^\beta_w\,. \tag{1.5}$$

Keeping γ fixed the matrix of the coefficients of the derivatives $p_{\alpha\beta\gamma}$ in the equations (1.4) and (1.5) has the form

$$
\left\| \begin{matrix}
a^{11} & 2a^{12} & a^{22} & 2a^{13} & 2a^{23} & a^{33} & b_\gamma \\
x^1_u & x^2_u & 0 & x^3_u & 0 & 0 & -\dfrac{\partial p_{1\gamma}}{\partial u} \\
x^1_w & x^2_w & 0 & x^3_w & 0 & 0 & -\dfrac{\partial p_{1\gamma}}{\partial w} \\
0 & x^1_u & x^2_u & 0 & x^3_u & 0 & \dfrac{\partial p_{2\gamma}}{\partial u} \\
0 & x^1_w & x^2_w & 0 & x^3_w & 0 & -\dfrac{\partial p_{2\gamma}}{\partial w} \\
0 & 0 & 0 & x^1_u & x^2_u & x^3_u & -\dfrac{\partial p_{3\gamma}}{\partial u} \\
0 & 0 & 0 & x^1_w & x^2_w & x^3_w & -\dfrac{\partial p_{3\gamma}}{\partial w}
\end{matrix} \right\| . \tag{1.6}
$$

The six rowed determinant in the upper left hand corner of (1.6) has the value

$$
-x^3_u a^{\alpha\beta} D_\alpha D_\beta ,
$$

where

$$
D_1 = x^2_u x^3_w - x^3_u x^2_w , \qquad D_2 = x^3_u x^1_w - x^1_u x^3_w , \qquad D_3 = x^1_u x^2_w - x^2_u x^1_w .
$$

Under the assumption that the initial strip (1.2) satisfies the condition $a^{\alpha\beta} D_\alpha D_\beta \neq 0$, we can solve the first six equations of (1.4) and (1.5) for the six derivatives $p_{\alpha\beta\gamma}$ (γ fixed). There is no loss of generality in assuming in particular that $x^3_u \neq 0$.

Now let us consider the characteristic case, i.e. the case in which the strip (1.2) satisfies the equation

$$
a^{\alpha\beta} D_\alpha D_\beta = 0. \tag{1.7}
$$

The six rowed determinant in the lower left hand corner of (1.6) vanishes identically. For if we multiply the first row by x^1_w, the second by $-x^1_u$, etc., and add, we find that the rows are linearly dependent. Hence, the equation (1.7) implies that the rank of the matrix, formed by taking the first six columns of (1.6), is less than six. Let us assume, for example, that the determinant

$$
\left| \begin{matrix}
x^1_u & 0 & x^3_u & 0 & 0 \\
x^1_w & 0 & x^3_w & 0 & 0 \\
0 & x^2_u & 0 & x^3_u & 0 \\
0 & x^2_w & 0 & x^3_w & 0 \\
0 & 0 & x^1_w & x^2_w & x^3_w
\end{matrix} \right| = x^3_w D_1 D_2 \neq 0. \tag{1.8}
$$

Then if the equations (1.4) and (1.5) are to be consistent, we have six conditions which must be satisfied, namely

$$(1.9)\qquad \begin{vmatrix} a^{11} & a^{22} & 2a^{13} & 2a^{23} & a^{33} & b_\gamma \\ x_u^1 & 0 & x_u^3 & 0 & 0 & -\dfrac{\partial p_{1\gamma}}{\partial u} \\ x_w^1 & 0 & x_w^3 & 0 & 0 & -\dfrac{\partial p_{1\gamma}}{\partial w} \\ 0 & x_u^2 & 0 & x_u^3 & 0 & -\dfrac{\partial p_{2\gamma}}{\partial u} \\ 0 & x_w^2 & 0 & x_w^3 & 0 & -\dfrac{\partial p_{2\gamma}}{\partial w} \\ 0 & 0 & x_w^1 & x_w^2 & x_w^3 & -\dfrac{\partial p_{3\gamma}}{\partial w} \end{vmatrix} = 0,$$

and

$$(1.10)\qquad x_w^\alpha \frac{\partial p_{\alpha\gamma}}{\partial u} - x_u^\alpha \frac{\partial p_{\alpha\gamma}}{\partial w} = 0.$$

A strip (1.2) which satisfies (1.7) and conditions of the type (1.8), (1.9), and (1.10) will be called a *two-dimensional characteristic strip.* If we know a solution, $z = z^*(x^\alpha)$, of the equation (1.1) then the quantities $a^{\alpha\beta}$ are known functions of x^α; let us denote them by $a^{*\alpha\beta}$. A surface $x^\alpha = \eta^\alpha(u, w)$ which satisfies the equation

$$(1.11)\qquad a^{*\alpha\beta}D_\alpha D_\beta = 0$$

will be called a *characteristic surface for the solution* z^*. If the non-parametric equation of the surface is $\Phi(x^1, x^2, x^3) = 0$ then (1.11) becomes

$$(1.12)\qquad a^{*\alpha\beta}\Phi_\alpha\Phi_\beta = 0.$$

2. The initial value problem for $F = 0$

Thus far we have been dealing with Cauchy's problem in general. In this section we shall formulate the Cauchy problem which will interest us in the present paper, i.e. we shall formulate conditions to be imposed upon the function F and upon the initial strip (1.2).

We propose the following initial value problem for the equation (1.1). Let the function F and the initial strip

$$(2.1)\qquad x^\alpha = \bar{x}^\alpha(s, w); \quad z = \bar{z}(s, w); \quad p_\alpha = \bar{p}_\alpha(s, w); \quad p_{\alpha\beta} = \bar{p}_{\alpha\beta}(s, w);$$

satisfy the following conditions:

(a) The functions (2.1) and their first derivatives with respect to s shall be

partially analytic with respect to w. The strip functions (2.1) shall satisfy the equation (1.1) and the strip conditions

$$\bar{z}_s = \bar{p}_\alpha \bar{x}_s^\alpha, \qquad \bar{z}_w = \bar{p}_\alpha \bar{x}_w^\alpha,$$
$$\frac{\partial \bar{p}_\alpha}{\partial s} = \bar{p}_{\alpha\beta} \bar{x}_s^\beta, \qquad \frac{\partial \bar{p}_\alpha}{\partial w} = \bar{p}_{\alpha\beta} \bar{x}_w^\beta. \tag{2.2}$$

(b) The function F shall be an analytic function of its arguments in a neighborhood of the set of values $\bar{x}^\alpha$, $\bar{z}$, $\bar{p}_\alpha$, $\bar{p}_{\alpha\beta}$.

(c) Along the strip (2.1) the following conditions shall be satisfied:

$$a^{\alpha\beta}\Delta_\alpha\Delta_\beta \neq 0, \tag{2.3}$$

$$A_{\alpha\beta}x_w^\alpha x_w^\beta < 0, \tag{2.4}$$

where the Δ_α are analogous to the D_α with u replaced by s; the $A_{\alpha\beta}$ denote the cofactors of $a^{\alpha\beta}$ in the determinant $a = | a^{\alpha\beta} |$.

The problem of determining a solution of (1.1) containing the strip (2.1) will be termed the initial value problem for $F = 0$. The remainder of §2 is devoted to interpreting the conditions (2.3) and (2.4) algebraically and geometrically in order that our problem can be compared to other Cauchy problems. In the first part of §3 we introduce quantities $\mathfrak{A}_{\alpha\beta}$ for future reference and then discuss the invariance of our problem under changes of independent variable x^α and changes of parameter s and w in (2.1). We then reduce our problem to a canonical form termed the initial value problem for $\bar{F} = 0$. In §4 under the assumption that we have a solution z^* of the initial value problem for $\bar{F} = 0$, we map the space of coördinates x^α onto a space of coördinates u, v, w in such a manner that two one-parameter family characteristic surfaces for z^* map into the planes u = constant and v = constant. We then form a system of partial differential equations Σ with independent variables u, v, w by selecting certain of the characteristic conditions in §1 which must be satisfied over u = constant and v = constant. In §5, with the aid of a uniqueness theorem in §6, we show that a partially analytic solution of the induced initial value problem for Σ is also a solution of the initial value problem for $\bar{F} = 0$. In §9, with the aid of an existence theorem in §8, we show the existence of a partially analytic solution of this induced initial value problem for Σ. Also in §9, with the aid of a uniqueness theorem in §8, we show the uniqueness of a solution z^* of the initial value problem for $\bar{F} = 0$ for which the mapping in §4 is partially analytic. In §7 we point out certain simplifications in case $F = 0$ is linear and obtain a slightly different uniqueness result.

We shall now examine the significance of the above conditions. The condition (2.3) insures that *the surface* S: $x^\alpha = \bar{x}^\alpha(s, w)$ *shall be non-singular and the rank of the matrix* $\| a \|$ *must be two or three,* if the condition (2.4) is to be satisfied.

Before proceeding further we observe an identity that will be of use to us several times. If we apply Lagrange's identity

$$(2.5) \qquad (a \times b)\cdot(c \times d) = (a\cdot c)(b\cdot d) - (a\cdot d)(b\cdot c),$$

(written in the notation of vector analysis) four times, we get

$$(2.6) \qquad \begin{vmatrix} a^{\alpha\beta}Y_\alpha \bar{Y}_\beta & a^{\alpha\beta}Y_\alpha \bar{Z}_\beta \\ a^{\alpha\beta}Z_\alpha \bar{Y}_\beta & a^{\alpha\beta}Z_\alpha \bar{Z}_\beta \end{vmatrix} \equiv \begin{vmatrix} a^{\alpha 2}Y_\alpha & a^{\alpha 3}Y_\alpha \\ a^{\alpha 2}Z_\alpha & a^{\alpha 3}Z_\alpha \end{vmatrix} \begin{vmatrix} \bar{Y}_2 & \bar{Y}_3 \\ \bar{Z}_2 & \bar{Z}_3 \end{vmatrix} + \cdots$$

$$\equiv (A_{\alpha 1}R^\alpha)\bar{R}^1 + \cdots \equiv A_{\alpha\beta}R^\alpha \bar{R}^\beta,$$

where

$$R^1 = Y_2Z_3 - Y_3Z_2, \qquad R^2 = Y_3Z_1 - Y_1Z_3, \qquad R^3 = Y_1Z_2 - Y_2Z_1,$$

and similarly for $\bar{R}^\alpha$.[9]

The condition (2.4) implies that the form (2.3) must be indefinite (semi-indefinite) for if the form (2.3) were definite (semi-definite) then the adjoint form (2.4) would be positive definite (positive semi-definite) and (2.4) could not be satisfied by a real vector x_w^α. The case of rank three is covered by a discussion of Dickson[10] and we give here a proof following the same lines which covers the case of rank two.

Let us assume that the form $\psi \equiv a^{\alpha\beta}Y_\alpha Y_\beta$ is positive semi-definite of rank two, i.e. $\psi \geqq 0$ for real Y_α. We can without loss of generality assume that $a^{11} \neq 0$, for if all the $a^{\alpha\alpha}$ were zero then at least one of the $a^{\alpha\beta}$ $(\alpha \neq \beta)$ would be different from zero and ψ would take on both positive and negative values. Since $\psi(1, 0, 0) = a^{11}$, we have $a^{11} > 0$. By taking $Y_\alpha = \bar{Y}_\alpha$, $Z_\alpha = \bar{Z}_\alpha$, $Z_1 = 1$, $Z_2 = Z_3 = 0$ in (2.6) we have

$$(2.7) \qquad a^{11}\psi(Y_1, Y_2, Y_3) \equiv L^2 + A_{33}(Y_2)^2 - 2A_{23}Y_2Y_3 + A_{22}(Y_3)^2,$$

where $L = a^{\alpha 1}Y_\alpha$. One of the quantities A_{22}, A_{33}, let us say $A_{33} \neq 0$, for if both were zero then the principle minors obtained by bordering a^{11} with one row and one column, and also with two rows and two columns would vanish and $\|a\|$ would be of rank one. Then take $Y_1 = -a^{12}$, $Y_2 = a^{11}$, $Y_3 = 0$ in (2.7) and it follows that $A_{33} > 0$. Multiply (2.7) by A_{33} and use $|a| = 0$, we get

$$a^{11}A_{33}\psi(Y_1, Y_2, Y_3) \equiv A_{33}L^2 + (A_{33}Y_2 - A_{23}Y_3)^2.$$

THEOREM I. *In a positive semi-definite form (rank 2) by a change of subscript we can make a^{11}, A_{33} positive. If a^{11} and A_{33} are positive then a form (rank 2) is positive semi-definite.* In a negative semi-definite form replace the above by $a^{11} < 0$ and $A_{33} > 0$.

---

[9] The frequent use of Lagrange's identity in what follows is the outgrowth of a remark by Dr. Tobias Dantzig.

[10] See L. E. Dickson, *Studies in the Theory of Numbers*, University of Chicago Press (1930), pp. 1–10.

The form $\Psi \equiv A_{\alpha\beta}R^\alpha R^\beta$ is of rank one if ψ is of rank two. Hence, if we apply (2.7) to Ψ, we see that the adjoint of a semi-definite form (rank 2) is a positive semi-definite form (rank 1). We sum up the above remarks by saying that (2.4) *implies that the equation* (1.1) *is hyperbolic for the initial data* (2.1).

Let us for the present think of

$$a^{\alpha\beta}\Phi_\alpha\Phi_\beta = 0 \tag{2.8}$$

as a conic in homogeneous point coordinates Φ_α and think of

$$x_w^\alpha \Phi_\alpha = 0 \tag{2.9}$$

as a line.

THEOREM II. *The inequality* (2.4) *is a necessary and sufficient condition that the line* (2.9) *cut the conic* (2.8) *in two real and distinct points* Φ_α.

To show that the condition is necessary let Y_α and Z_α be two real and distinct points on the line (2.9) which do not lie on the conic (2.8). Then the equation

$$a^{\alpha\beta}(Y_\alpha + \lambda Z_\alpha)(Y_\beta + \lambda Z_\beta) \equiv a^{\alpha\beta}Y_\alpha Y_\beta + 2\lambda a^{\alpha\beta}Y_\alpha Z_\beta + \lambda^2 a^{\alpha\beta}Z_\alpha Z_\beta = 0 \tag{2.10}$$

must have two real and distinct roots λ. Since the R^α are proportional to the x_w^α it then follows from (2.6) that the condition (2.4) is necessary. In order to be able to reverse the argument there must exist two points on the line but not on the conic. If every point on the line (2.9) satisfies (2.8) then pick two such points Y_α and Z_α. Equation (2.10) must be satisfied for all values of λ, i.e. the left member of (2.6) must vanish. Hence, the left member of (2.4) vanishes also, contrary to hypothesis.

We divide the discussion into two cases according to the rank of the matrix $\| a \|$.

RANK 3: The equation (2.8) can be regarded as the tangential equation of the *characteristic cone* having the corresponding point P of the surface S as its vertex.[11] Condition (2.3) *then implies that the tangent plane to* S *at* P *is not tangent to the characteristic cone with vertex at* P. Also (2.9) can be regarded as the tangential equation of the vector x_w^α which is a vector at P. From the above result on conics we see that the condition (2.4) can be regarded as requiring that through the vector x_w^α at P there pass two real planes tangent to the characteristic cone with vertex at P. *In other words* (2.4) *implies that the vector* x_w^α *at* P *lies outside the characteristic cone with vertex at* P.

RANK 2: In this case the characteristic cone becomes a pair of real intersecting lines through P. *Condition* (2.3) *then implies that the tangent plane to* S *at* P *does not contain either of the lines which constitute the characteristic cone with vertex at* P. The condition (2.4) implies that these same lines together with the vector x_w^α at P determine two distinct planes. *In other words* (2.4) *implies that the vector* x_w^α *at* P *does not lie in the plane of the characteristic cone with vertex at* P.

[11] Cf. *Loc. cit.* Hadamard (1923), p. 21.

We give three examples which indicate the scope of the initial value problem for $F = 0$.

$$\text{(A)} \begin{cases} F \equiv p_{11} + p_{22} - p_{33} = 0, & a^{\alpha\beta}\Delta_\alpha\Delta_\beta = -1, \\ S: x^3 = 0, \quad x^1 = s, \quad x^2 = w, & A_{\alpha\beta}x_w^\alpha x_w^\beta = -1, \end{cases}$$

$$\text{(B)} \begin{cases} F \equiv p_{11} + p_{22} - p_{33} = 0, & a^{\alpha\beta}\Delta_\alpha\Delta_\beta = 1, \\ S: p^1 = 0, \quad x^3 = s, \quad x^2 = w, & A_{\alpha\beta}x_w^\alpha x_w^\beta = -1, \end{cases}$$

$$\text{(C)} \begin{cases} F \equiv p_{11} - p_{22} + p_3 = 0, & a^{\alpha\beta}\Delta_\alpha\Delta_\beta = 1, \\ S: x^1 = 0, \quad x^2 = s, \quad x^3 = w, & A_{\alpha\beta}x_w^\alpha x_w^\beta = -1. \end{cases}$$

In example (A) the characteristic cone with vertex at a point of S is $(x^1 - x_0^1)^2 + (x^2 - x_0^2)^2 - (x^3)^2 = 0$ and the vector $x_w^\alpha \sim 0, 1, 0$ lies outside this cone. We note that in (A) s and w could be interchanged. In (B) the characteristic cone is $(x^1)^2 + (x^2 - x_0^2)^2 - (x^3 - x_0^3)^2 = 0$ and $x_w^\alpha \sim 0, 1, 0$. Here x^1 and x^2 could be interchanged but not s and w. In (C) the characteristic cone consists of $(x^1)^2 - (x^2 - x_0^2)^2 = 0$, $x^3 = x_0^3$, a pair of lines. The vector $x_w^\alpha \sim 0, 0, 1$ is perpendicular to the plane of these lines. In (C) the surface S could have been $x^1 = s$, $x^2 = 0$, $x^3 = w$.

3. The initial value problem for $\bar{F} = 0$

We return to the consideration of the conic (2.8) and the line (2.9) together with the line

$$Q^\alpha\Phi_\alpha = 0. \tag{3.1}$$

The condition that the lines (2.9) and (3.1) intersect in a point on the conic (2.8) is given by

$$a^{\alpha\beta}Y_\alpha Y_\beta \equiv \mathfrak{A}_{\alpha\beta}Q^\alpha Q^\beta = 0, \tag{3.2}$$

where

$$Y_1 = Q^2x_w^3 - Q^3x_w^2, \qquad Y_2 = Q^3x_w^1 - Q^1x_w^3, \qquad Y_3 = Q^1x_w^2 - Q^2x_w^1.$$

The quantities $\mathfrak{A}_{\alpha\beta}$ are defined by (3.2). Recalling the manner in which we obtained (1.7) from (1.6) we get the following explicit formulae which occur often in our work

$$\mathfrak{A}_{11} = \begin{vmatrix} a^{22} & 2a^{23} & a^{33} \\ x_w^2 & x_w^3 & 0 \\ 0 & x_w^2 & x_w^3 \end{vmatrix}, \qquad \mathfrak{A}_{22} = \begin{vmatrix} a^{11} & 2a^{13} & a^{33} \\ x_w^1 & x_w^3 & 0 \\ 0 & x_w^1 & x_w^3 \end{vmatrix},$$

A simpler method for obtaining $\mathfrak{A}_{11}$ for present purposes is to put $Q^1 = 1$, $Q^2 = Q^3 = 0$ in (3.2). Then $Y_1 = 0$, $Y_2 = -x_w^3$, $Y_3 = x_w^2$ and $\mathfrak{A}_{11} = a^{\alpha\beta}Y_\alpha Y_\beta$. If we put $Z_1 = x_w^3$, $Z_2 = 0$, $Z_3 = -x_w^1$ and $W_1 = -x_w^2$, $W_2 = x_w^1$, $W_3 = 0$ then $\mathfrak{A}_{22} = a^{\alpha\beta}Z_\alpha Z_\beta$ and $\mathfrak{A}_{33} = a^{\alpha\beta}W_\alpha W_\beta$. Also $\mathfrak{A}_{12} = a^{\alpha\beta}Y_\alpha Z_\beta$, $\mathfrak{A}_{13} =$

$a^{\alpha\beta}Y_\alpha W_\beta$ and $\mathfrak{A}_{23} = a^{\alpha\beta}Z_\alpha W_\beta$. Then $R^1 = \overset{3}{x}_w \overset{1}{x}_w$, $R^2 = \overset{3}{x}_w \overset{2}{x}_w$, $R^3 = (\overset{3}{x}_w)^2$ and (2.6) yields

$$\mathfrak{A}_{11}\mathfrak{A}_{22} - (\mathfrak{A}_{12})^2 \equiv (\overset{3}{x}_w)^2 A_{\alpha\beta} x^\alpha_w x^\beta_w . \tag{3.3}$$

A similar calculation yields

$$-(\mathfrak{A}_{11}\mathfrak{A}_{23} - \mathfrak{A}_{13}\mathfrak{A}_{12}) \equiv \overset{3}{x}_w \overset{2}{x}_w A_{\alpha\beta} x^\alpha_w x^\beta_w .$$

These relations together with four analogous ones show that the adjoint to $\| \mathfrak{A}_{\alpha\beta} \|$ is of rank one. Hence, $\| \mathfrak{A}_{\alpha\beta} \|$ is of rank two. From (3.3) and (2.4) it follows that each principal minor of $| \mathfrak{A}_{\alpha\beta} |$ is negative or zero. Hence from Theorem I it follows that the form (3.2) quadratic in the Q^α, is semi-indefinite.

THEOREM III. *The form* (3.2) *factors into two real and distinct forms linear in the* Q^α. Geometrically this means that the family of lines (3.1) satisfying (3.2), i.e. the family of lines through the intersection of (2.8) and (2.9) is composed of two distinct real pencils.

Let us investigate the effect of a change of coördinates x^α upon our boundary value problem for $F = 0$. If we require that F transform as a scalar under the non-singular analytic transformation

$$\bar{x}^\alpha = \bar{x}^\alpha(x^1, x^2, x^3), \tag{3.4}$$

then the quantities $a^{\alpha\beta}$ must constitute the components of a contravariant tensor of second order, i.e.

$$\bar{a}^{\alpha\beta} = a^{\gamma\delta} \frac{\partial \bar{x}^\alpha}{\partial x^\gamma} \frac{\partial \bar{x}^\beta}{\partial x^\delta}.$$

It follows immediately that $\bar{a} = a \,| \partial\bar{x}/\partial x |^2$ and it is easily seen that[12]

$$\bar{A}_{\alpha\beta} = A_{\gamma\delta} \frac{\partial x^\gamma}{\partial \bar{x}^\alpha} \frac{\partial x^\delta}{\partial \bar{x}^\beta} \left| \frac{\partial \bar{x}}{\partial x} \right|^2 .$$

The quantities x^α_w are the components of a contravariant vector, i.e.

$$\bar{x}^\alpha_w = x^\beta_w \frac{\partial \bar{x}^\alpha}{\partial x^\beta},$$

and making use of (2.5) we have

$$\bar{\Delta}_1 = \begin{vmatrix} x^\beta_s \dfrac{\partial \bar{x}^2}{\partial x^\beta} & x^\beta_s \dfrac{\partial \bar{x}^3}{\partial x^\beta} \\ x^\beta_w \dfrac{\partial \bar{x}^2}{\partial x^\beta} & x^\beta_w \dfrac{\partial \bar{x}^3}{\partial x^\beta} \end{vmatrix} \equiv \begin{vmatrix} \dfrac{\partial \bar{x}^2}{\partial x^2} & \dfrac{\partial \bar{x}^2}{\partial x^3} \\ \dfrac{\partial \bar{x}^3}{\partial x^2} & \dfrac{\partial \bar{x}^3}{\partial x^3} \end{vmatrix} \cdot \begin{vmatrix} \overset{2}{x}_s & \overset{3}{x}_s \\ \overset{2}{x}_w & \overset{3}{x}_w \end{vmatrix} + \cdots$$

$$= \Delta_1 \frac{\partial x^1}{\partial \bar{x}^1} \left| \frac{\partial \bar{x}}{\partial x} \right| + \cdots \equiv \Delta_\alpha \frac{\partial x^\alpha}{\partial \bar{x}^1} \left| \frac{\partial \bar{x}}{\partial x} \right|$$

[12] A proof that will hold even in case $\| a \|$ is of rank two can be modeled along the lines of L. P. Eisenhart, *Riemannian Geometry*, Princeton University Press (1926), p. 14. A direct proof, similar to that given below for the law of transformation of the Δ_α, could be given.

or

$$\bar{\Delta}_\alpha = \Delta_\beta \frac{\partial x^\beta}{\partial \bar{x}^\alpha} \left| \frac{\partial \bar{x}}{\partial x} \right|.$$

In the above we have made use of the fact that the quantities $\partial x^\alpha/\partial \bar{x}^1$ satisfy the system

$$\frac{\partial \bar{x}^\beta}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial \bar{x}^1} = \delta_1^\beta.$$

Hence the laws of transformation for the quantities appearing in (2.3) and (2.4) are given by

$$\bar{a}^{\alpha\beta}\bar{\Delta}_\alpha\bar{\Delta}_\beta = a^{\alpha\beta}\Delta_\alpha\Delta_\beta \left| \frac{\partial \bar{x}}{\partial x} \right|^2, \tag{3.5}$$

$$\bar{A}_{\alpha\beta}\bar{x}_w^\alpha\bar{x}_w^\beta = A_{\alpha\beta}x_w^\alpha x_w^\beta \left| \frac{\partial \bar{x}}{\partial x} \right|^2.$$

Thus *the conditions* (2.3) *and* (2.4) *are invariant under non-singular transformations* (3.4).

Equation (3.5) can be written in the form

$$\bar{\mathfrak{A}}_{\alpha\beta}\bar{x}_s^\alpha\bar{x}_s^\beta = \mathfrak{A}_{\alpha\beta}x_s^\alpha x_s^\beta \left| \frac{\partial \bar{x}}{\partial x} \right|^2.$$

Make use of the law of transformation for the vector x_s^α and the quotient law of tensors, we get

$$\bar{\mathfrak{A}}_{\alpha\beta} = \mathfrak{A}_{\gamma\delta} \frac{\partial x^\gamma}{\partial \bar{x}^\alpha} \frac{\partial x^\delta}{\partial \bar{x}^\beta} \left| \frac{\partial \bar{x}}{\partial x} \right|^2.$$

In view of (2.3) and (2.4) at least one of the quantities $\mathfrak{A}_{\alpha\beta}$ and at least one of the quantities x_w^α must be different from zero at each point of the strip (2.1). Then choose the transformation (3.4) or more particularly the derivatives $\partial x^\alpha/\partial \bar{x}^1$ so that $\bar{\mathfrak{A}}_{11} \neq 0$ and the derivatives $\partial x^\alpha/\partial \bar{x}^2$ so that $\bar{\mathfrak{A}}_{22} \neq 0$. If necessary alter this choice to make the determinant

$$\begin{vmatrix} \dfrac{\partial x^1}{\partial \bar{x}^1} & \dfrac{\partial x^1}{\partial \bar{x}^2} & x_w^1 \\ \dfrac{\partial x^2}{\partial \bar{x}^1} & \dfrac{\partial x^2}{\partial \bar{x}^2} & x_w^2 \\ \dfrac{\partial x^3}{\partial \bar{x}^1} & \dfrac{\partial x^3}{\partial \bar{x}^2} & x_w^3 \end{vmatrix}$$

different from zero. The value of $\bar{x}_w^3$ obtained by solving the system of three equations

$$\frac{\partial x^\alpha}{\partial \bar{x}^\beta} \bar{x}_w^\beta = x_w^\alpha$$

will be different from zero.

Suppose that change of variable x^α of the above type has been made; we will now make a change of parameter that will make $\bar{x}^3 = w$. Since $\bar{x}^3_w \neq 0$, the change of parameter

$$\bar{s} = s, \qquad \bar{w} = \bar{x}^3(s, w), \tag{3.6}$$

has an inverse of the form

$$s = \bar{s}, \qquad w = w(\bar{s}, \bar{w}),$$

where w is partially analytic with respect to $\bar{w}$. From (3.6) we have

$$\frac{\partial x^\alpha}{\partial s} = \frac{\partial x^\alpha}{\partial \bar{s}} + \frac{\partial x^\alpha}{\partial \bar{w}} \frac{\partial \bar{w}}{\partial s}, \qquad \frac{\partial x^\alpha}{\partial w} = \frac{\partial x^\alpha}{\partial \bar{w}} \frac{\partial \bar{w}}{\partial w},$$

and it is easily seen that each of the conditions (2.3), (2.4), $\mathfrak{A}_{11} \neq 0$, $\mathfrak{A}_{22} \neq 0$, remain unchanged under the above change of parameter.

After the change of variable (3.4) followed by the change of parameter (3.6) has been made we shall refer to our initial value problem as the problem for $\bar{F} = 0$. The initial value problem for $\bar{F} = 0$ satisfies in addition to conditions (a), (b), (c) the conditions

$$(c') \quad \begin{cases} (3.7) \quad \bar{x}^3 = w, \\ (3.8) \quad \mathfrak{A}_{11}(x^1_s)^2 + 2\mathfrak{A}_{12}x^1_s x^2_s + \mathfrak{A}_{22}(x^2_s)^2 \neq 0, \\ (3.9) \quad \mathfrak{A}_{11}\mathfrak{A}_{22} - (\mathfrak{A}_{12})^2 < 0, \\ (3.10) \quad \mathfrak{A}_{11} \neq 0, \quad \mathfrak{A}_{22} \neq 0. \end{cases}$$

4. The initial value problem for Σ

In this section we assume that we have a solution of our boundary value problem for $\bar{F} = 0$, namely a function $z = z^*(x^1, x^2, x^3)$ which satisfies (1.1) and contains the strip (2.1). If we look for characteristic surfaces for z^* in the form $\Phi(x^\alpha) \equiv x^1 - \varphi(x^2, x^3) = 0$ then the partial differential equation (1.12) becomes

$$\Omega \equiv a^{*22}(\varphi_2)^2 + 2a^{*23}\varphi_2\varphi_3 + a^{*33}(\varphi_3)^2 - 2a^{*12}\varphi_2 - 2a^{*13}\varphi_3 + a^{*11} = 0. \tag{4.1}$$

Cauchy's method of integrating the equation (4.1) employs the one-dimensional characteristic strips which are solutions of the system of ordinary equations

$$\frac{dx^a}{d\nu} = \frac{\partial \Omega}{\partial \varphi_a}, \qquad \frac{dx^1}{d\nu} = \varphi_a \frac{\partial \Omega}{\partial \varphi_a}, \qquad \frac{d\varphi_a}{d\nu} = -\frac{\partial \Omega}{\partial x^a} - \frac{\partial \Omega}{\partial x^1}\varphi_a \qquad (a = 2, 3). \tag{4.2}$$

Then take a solution of the system (4.2) depending on five arbitrary constants x_0^α, φ_{a0}, namely

$$(4.3)\qquad \begin{aligned} x^\alpha &= x^\alpha(\nu \mid x_0^\alpha \mid \varphi_{a0}), \\ \varphi_a &= \varphi_a(\nu \mid x_0^\alpha \mid \varphi_{a0}), \end{aligned}$$

which reduces to x_0^α, φ_{a0} for $\nu = 0$.

The Cauchy problem which will interest us is to pass a solution of (4.1) through each of the curves C^w (s = constant) lying on the surface S. The initial strips are obtained by solving the two equations (4.1) and

$$(4.4)\qquad -x_w^1 + x_w^2\varphi_2 + x_w^3\varphi_3 = 0$$

for φ_2 and φ_3 for each value of s. Take account of (3.7), multiply (4.1) by $(x_w^3)^2$, use (4.4); one obtains

$$(4.5)\qquad \mathfrak{A}_{11}(\varphi_2)^2 + 2\mathfrak{A}_{12}\varphi_2 + \mathfrak{A}_{22} = 0.$$

On account of (3.9) and (3.10) the two solutions $\bar{\varphi}_2(s, w)$, $\bar{\bar{\varphi}}_2(s, w)$ of (4.5) will be real and distinct. Thus when we replace the arbitrary constants in (4.3) by $\bar{x}^\alpha(s, w)$, $\bar{\varphi}_a(s, w)$; $\bar{x}^\alpha(s, w)$, $\bar{\bar{\varphi}}_a(s, w)$, ($s$ = constant) we obtain two surfaces satisfying (4.1) through each of the curves C^w.

If instead of replacing the arbitrary constants in (4.3) by functions of w alone, we replace them by the functions $\bar{x}^\alpha(s, w)$, $\bar{\varphi}_a(s, w)$, also by $\bar{x}^\alpha(s, w)$, $\bar{\bar{\varphi}}_a(s, w)$ we obtain two sets of functions

$$(4.6)\qquad \text{(a) } x^\alpha = X^\alpha(\nu, s, w), \qquad \text{(b) } x^\alpha = \Xi^\alpha(\nu, s, w)$$

each of which gives a mapping of the portion of the x^α space in the neighborhood of S upon the space with coördinates ν, s, w. We have merely to show that the Jacobian of (4.6) is different from zero at each point of S. On account of (4.2) at a point of S ($\nu = 0$) the Jacobian of (4.6a) is

$$(4.7)\qquad \begin{vmatrix} 2a^{\alpha\beta}Y_\alpha\delta_\beta^1 & \bar{x}_s^1 & \bar{x}_w^1 \\ 2a^{\alpha\beta}Y_\alpha\delta_\beta^2 & \bar{x}_s^2 & \bar{x}_w^2 \\ 2a^{\alpha\beta}Y_\alpha\delta_\beta^3 & \bar{x}_s^3 & \bar{x}_w^3 \end{vmatrix} \equiv 2a^{\alpha\beta}Y_\alpha Z_\beta \neq 0,$$

where $Y_1 = -1$, $Y_a = \bar{\varphi}_a$; $Z_\alpha = \Delta_\alpha$. By (4.1) Y_α lies on the conic (2.8) and by (2.3) Z_α lies off the same conic. In view of (4.4) both Y_α and Z_α lie on the line (2.9). Hence (4.7) must be satisfied or (2.10) would have two roots $\lambda = 0$ and (2.9) would be tangent to (2.8).

For $\nu = 0$, as a function of s and w, the determinant

$$(4.8)\qquad \begin{vmatrix} X_\nu^1 & X_\nu^2 & X_\nu^3 \\ \Xi_\nu^1 & \Xi_\nu^2 & \Xi_\nu^3 \\ \bar{x}_w^1 & \bar{x}_w^2 & \bar{x}_w^3 \end{vmatrix} = \begin{vmatrix} 2a^{\alpha\beta}Y_\alpha\delta_\beta^1 & 2a^{\alpha\beta}Y_\alpha\delta_\beta^2 & 2a^{\alpha\beta}Y_\alpha\delta_\beta^3 \\ 2a^{\alpha\beta}Z_\alpha\delta_\beta^1 & 2a^{\alpha\beta}Z_\alpha\delta_\beta^2 & 2a^{\alpha\beta}Z_\alpha\delta_\beta^3 \\ \bar{x}_w^1 & \bar{x}_w^2 & \bar{x}_w^3 \end{vmatrix},$$

where $Z_\alpha \sim -1, \bar{\varphi}_2, \bar{\varphi}_3$. If we write $R^1 = \bar{\varphi}_2\bar{\varphi}_3 - \bar{\varphi}_3\bar{\varphi}_2$, $R^2 = \bar{\varphi}_3 - \bar{\varphi}_3$, $R^3 = \bar{\varphi}_2 - \bar{\varphi}_2$, then on account of identities of the type (2.6) the quantity (4.8) becomes $4A_{\alpha\beta}R^\alpha \bar{x}^\beta_w$. But by (4.4) the R^α are proportional to the $\bar{x}^\alpha_w$ and hence by (2.4) the quantity (4.8) is different from zero over S ($\nu = 0$).

Consider the change of variable

$$\text{(4.9)} \qquad \text{(a)} \quad \begin{aligned} \bar{\nu} &= \nu, \\ \bar{s} &= s, \\ \bar{w} &= X^3(\nu, s, w), \end{aligned} \qquad \text{(b)} \quad \begin{aligned} \tilde{\nu} &= \nu, \\ \tilde{s} &= s, \\ \tilde{w} &= \Xi^3(\nu, s, w). \end{aligned}$$

Since (4.9) is the identity transformation for $\nu = 0$, the initial value problem for $\bar{F} = 0$ will not be affected by this change of parameter. For the same reason the set of transformed functions (4.6), namely

$$\text{(4.10)} \qquad \begin{aligned} \text{(a)} \quad x^i &= \bar{X}^i(\bar{\nu}, \bar{s}, \bar{w}), \\ x^3 &= \bar{w}, \end{aligned} \qquad \begin{aligned} \text{(b)} \quad x^i &= \tilde{\Xi}^i(\tilde{\nu}, \tilde{s}, \tilde{w}), \\ x^3 &= \tilde{w}, \end{aligned} \qquad (i = 1, 2)$$

will reduce to $\bar{x}^\alpha(s, w)$ if we put $\bar{\nu} = 0$ or $\tilde{\nu} = 0$ as the case may be, and put $s = \bar{s} = \tilde{s}$, $w = \bar{w} = \tilde{w}$. From the form of (4.9) it is easily seen that the surfaces obtained by putting $\bar{s}$ = constant or $\tilde{s}$ = constant in (4.10) are characteristic. On account of the nonvanishing of (4.8) and the form of (4.9) it follows that for $\bar{v} = \tilde{\nu} = 0$, $s = \bar{s} = \tilde{s}$, $w = \bar{w} = \tilde{w}$, the quantity

$$\text{(4.11)} \qquad \begin{vmatrix} \bar{X}^1_{\bar{\nu}} & \bar{X}^2_{\bar{\nu}} & 0 \\ \tilde{\Xi}^1_{\tilde{\nu}} & \tilde{\Xi}^2_{\tilde{\nu}} & 0 \\ \bar{X}^1_{\bar{w}} & \bar{X}^2_{\bar{w}} & \bar{X}^3_{\bar{w}} \end{vmatrix} \neq 0.$$

Let us identify $w, \bar{w}, \tilde{w}$, put $\bar{s} = \sqrt{2}\, u$, $\tilde{s} = -\sqrt{2}\, v$ and consider

$$\text{(4.12)} \qquad \bar{X}^i(\bar{\nu}, \sqrt{2}\, u, w) - \tilde{\Xi}^i(\tilde{\nu}, -\sqrt{2}\, v, w) = 0 \qquad (i = 1, 2)$$

as equations for $\bar{\nu}$ and $\tilde{\nu}$ in terms of u, v, w. For $u + v = 0$ the equations (4.12) possess the solution $\bar{\nu} = \tilde{\nu} = 0$, and by (4.11) the Jacobian of the left members of (4.12) with respect to $\bar{\nu}, \tilde{\nu}$ is different from zero for $u + v = 0$. Hence, (4.12) possesses a unique solution $\bar{\nu} = \bar{\nu}(u, v, w)$, $\tilde{\nu} = \tilde{\nu}(u, v, w)$ which reduces to $\bar{\nu} = \tilde{\nu} = 0$ for $u + v = 0$. Differentiating (4.12) we find that the Jacobian

$$\frac{\partial(\bar{\nu}\bar{s}\bar{w})}{\partial(uvw)} = -\sqrt{2}\, \frac{\partial \bar{\nu}}{\partial v} = -2\, \frac{\dfrac{\partial(\tilde{\Xi}^1 \ \tilde{\Xi}^2)}{\partial(\tilde{\nu} \ \tilde{s})}}{\begin{vmatrix} \bar{X}^1_{\bar{\nu}} & \bar{X}^2_{\bar{\nu}} \\ \tilde{\Xi}^1_{\tilde{\nu}} & \tilde{\Xi}^2_{\tilde{\nu}} \end{vmatrix}} \neq 0,$$

for $u + v = 0$. Hence, from either of the mappings (4.10) we obtain

$$\text{(4.13)} \qquad x^i = x^i(u, v, w), \qquad x^3 = w \qquad (i = 1, 2),$$

which maps the portion of x^α space in the neighborhood of S upon the space of coördinates u, v, w. If we regard the u, v, w space as Euclidean with rectangular Cartesian coördinates then the planes w = constant map into the surfaces x^3 = constant. The plane $u + v = 0$ maps into S, i.e. (4.13) reduces to $\bar{x}^\alpha(s, w)$ if we put $\sqrt{2}\, u = -\sqrt{2}\, v = s$. The planes u = constant and v = constant map into characteristic surfaces. The variable s measures distance along the lines $u + v = 0$, w = constant.

Upon replacing the x^α in the z^* and its derivatives by means of equations (4.13) we obtain a set of functions

$$(4.14) \qquad x^\alpha(u, v, w); \qquad z(u, v, w); \qquad p_\alpha(u, v, w); \qquad p_{\alpha\beta}(u, v, w)$$

which reduces to (2.1) for $u + v = 0$. A repetition of the argument used in §1 shows that the functions (4.14) for u = constant constitute a two-dimensional characteristic strip; similarly for v = constant. In fact the functions (4.14) must satisfy the equations

$$(4.15) \qquad \begin{aligned} \mathfrak{A}_{11}(x^1_u)^2 + 2\mathfrak{A}_{12}x^1_u x^2_u + \mathfrak{A}_{22}(x^2_u)^2 &= 0, \\ \mathfrak{A}_{11}(x^1_v)^2 + 2\mathfrak{A}_{12}x^1_v x^2_v + \mathfrak{A}_{22}(x^2_v)^2 &= 0. \end{aligned}$$

In view of (3.9) and (3.10) the two roots ρ and σ of the quadratic equation (4.15) are real, distinct and different from zero in the neighborhood of $S(u + v = 0)$. Since in (4.13) $x^1_u x^2_v - x^2_u x^1_v \neq 0$, the equations (4.15) can be replaced by the first two equations of the system (4.16). The remaining equations of (4.16), which we shall refer to as the system Σ, are some of the conditions which express the fact that (4.14) for u = constant or v = constant constitute a two-dimensional characteristic strip.

$$(4.16) \qquad \begin{aligned}
&(a) \quad \begin{cases} x^2_u - \rho\, x^1_u = 0, \\ x^2_v - \sigma\, x^1_v = 0, \\ x^3_u = 0, \end{cases} \\
&(b) \quad \begin{cases} z_u = p_\alpha x^\alpha_u, \\ \dfrac{\partial p_\alpha}{\partial u} = p_{\alpha\beta} x^\alpha_u, \end{cases} \\
&(c) \quad \left[\mathfrak{A}_{22}\, \rho \frac{\partial p_{1\gamma}}{\partial u} + \mathfrak{A}_{11} \frac{\partial p_{2\gamma}}{\partial u} \right] x^3_w + E_\gamma x^2_u = 0, \\
&(d) \quad x^\alpha_w \frac{\partial p_{a\alpha}}{\partial u} - \frac{\partial p_{a\alpha}}{\partial w} x^\alpha_u = 0 \qquad (a = 2, 3), \\
&(e) \quad \left[\mathfrak{A}_{22}\, \sigma \frac{\partial p_{11}}{\partial v} + \mathfrak{A}_{11} \frac{\partial p_{12}}{\partial v} \right] x^3_w + E_1 x^2_v = 0.
\end{aligned}$$

The equations (4.16c) are obtained by expanding (1.9) and making use of (4.16a). Thus the quantities E_γ depend on $a^{\alpha\beta}$, b_γ, $\partial p_{\alpha\gamma}/\partial w$, x^α_w. Conversely,

if $x_u^1 \neq 0$ the equations (4.16c) and (4.16a) imply (1.9), and we shall find in computation that it is often more convenient to use the form (1.9), rather than (4.16c).

We shall refer to the problem of finding a solution (4.14) of the system Σ which reduces to (2.1) for $u + v = 0$ as the initial value problem for Σ. Evidently a solution of the initial value problem for $\bar{F} = 0$ is also a solution of the initial value problem for Σ.

5. Equivalence

This section will be devoted to showing that a solution (4.14) of the initial value problem for Σ, which is partially analytic with respect to w, is also a solution of the initial value problem for $\bar{F} = 0$. To do this we must show that:

(1) The variables x^α can be introduced as independent variables in place of u, v, w.

(2) The functions (4.14) satisfy (1.1).

(3) $\dfrac{\partial z}{\partial x^\alpha} = p_\alpha, \quad \dfrac{\partial^2 z}{\partial x^\alpha \partial x^\beta} = p_{\alpha\beta}.$

First consider the one equation $x_u^3 = 0$ with initial conditions $\bar{x}^3 = w$. Evidently the function $x^3 = w$ satisfies the above equation and initial conditions. The uniqueness theorem in §6 shows that the solution is unique. Thus the function $x^3(u, v, w)$ is independent of u and v.

In order to show that, in (4.14), we can introduce the variables x^α as independent variables in place of u, v, w we must show that the Jacobian

$$\begin{vmatrix} x_u^1 & x_u^2 & 0 \\ x_v^1 & x_v^2 & 0 \\ x_w^1 & x_w^2 & x_w^3 \end{vmatrix} \neq 0$$

over the initial plane $u + v = 0$. The quantity

$$x_w^3(x_u^1 x_v^2 - x_u^2 x_v^1) = x_w^3 x_u^1 x_v^1(\sigma - \rho)$$

is different from zero provided we can show that x_u^1 and x_v^1 do not vanish initially. Suppose that x_u^1 does vanish over $u + v = 0$; then from the form of (4.16a) it follows that x_u^2 also vanishes. Since s is the arc length along the line $u + v = 0$, $w =$ constant, and $u = (1/\sqrt{2})s$ and $v = (-1/\sqrt{2})s$, we get

$$\frac{\partial x^i}{\partial s} = \left(\frac{\partial x^i}{\partial u} - \frac{\partial x^i}{\partial v}\right)\frac{1}{\sqrt{2}} = \frac{-1}{\sqrt{2}}\frac{\partial x^i}{\partial v} \qquad (i = 1, 2).$$

Hence

$$\frac{\partial x^2}{\partial s} - \sigma\frac{\partial x^1}{\partial s} = \frac{-1}{\sqrt{2}}\left(\frac{\partial x^2}{\partial v} - \sigma\frac{\partial x^1}{\partial v}\right) = 0,$$

and in view of (3.8) our assumption that x_u^1 vanishes leads to a contradiction.

Similarly x_v^1 must be different from zero over $u + v = 0$.

Let us adopt the notation

$$U = z_u - p_\alpha x_u^\alpha, \qquad V = z_v - p_\alpha x_v^\alpha, \qquad W = z_w - p_\alpha x_w^\alpha,$$

$$U_\alpha = \frac{\partial p_\alpha}{\partial u} - p_{\alpha\beta} x_u^\beta, \qquad V_\alpha = \frac{\partial p_\alpha}{\partial v} - p_{\alpha\beta} x_v^\beta, \qquad W_\alpha = \frac{\partial p_\alpha}{\partial w} - p_{\alpha\beta} x_w^\beta.$$

From the differential equations (4.16b) we have $U \equiv U_\alpha \equiv 0$. Making use of the fact that the solution (4.14) of the system Σ has continuous cross derivatives, we have

$$\frac{\partial W_\alpha}{\partial u} - \frac{\partial U_\alpha}{\partial w} = \frac{\partial p_{\alpha\beta}}{\partial w} x_u^\beta - \frac{\partial p_{\alpha\beta}}{\partial u} x_w^\beta. \tag{5.1}$$

Consequently the equations (4.16d) imply that

$$\frac{\partial W_2}{\partial u} = \frac{\partial W_3}{\partial u} = 0. \tag{5.2}$$

In consequence of condition (c′) the ratio $x_s^2 : x_s^1$ is different from $x_u^2 : x_u^1$ initially. Hence the determinant

$$\begin{vmatrix} x_u^1 & x_u^2 & 0 \\ x_s^1 & x_s^2 & 0 \\ x_w^1 & x_w^2 & x_w^2 \end{vmatrix} \neq 0$$

over $u + v = 0$, and it follows from (2.2) and (4.16b) that $\partial z/\partial x^\alpha = p_\alpha$, $\partial p_\alpha/\partial x^\beta = p_{\alpha\beta}$ over $u + v = 0$. In other words

$$V = V_\alpha = W = W_\alpha = 0 \quad \text{over} \quad u + v = 0, \tag{5.3}$$

which together with (5.2) implies that $W_2 \equiv W_3 \equiv 0$.

Multiply the last columns of the determinants (1.9) by x_u^γ respectively and add the resulting equations. Then in the determinant in the left member of the resulting equation, multiply the first five columns by certain factors and add to the last column and we obtain

$$\begin{vmatrix} a^{11} & a^{22} & 2a^{13} & 2a^{23} & a^{33} & \sum' a^{\alpha\beta} \dfrac{\partial p_{\alpha\beta}}{\partial u} + b_\gamma x_u^\gamma \\ x_u^1 & 0 & 0 & 0 & 0 & -\dfrac{\partial p_{12}}{\partial u} x_u^2 \\ x_w^1 & 0 & x_w^3 & 0 & 0 & -\dfrac{\partial p_{12}}{\partial u} x_w^2 - \dfrac{\partial W_1}{\partial u} \\ 0 & x_u^2 & 0 & 0 & 0 & -\dfrac{\partial p_{12}}{\partial u} x_u^1 \\ 0 & x_w^2 & 0 & x_w^3 & 0 & -\dfrac{\partial p_{12}}{\partial u} x_w^1 \\ 0 & 0 & x_w^1 & x_w^2 & x_w^3 & 0 \end{vmatrix} = 0,$$

where $\sum'$ denotes the sum indicated except that the term containing a^{12} is lacking. Expanding the above determinant we have

$$x_w^3 \frac{\partial p_{12}}{\partial u} a^{\alpha\beta} D_\alpha D_\beta + x_w^3 x_u^1 x_u^2 \left\{ (x_w^3)^2 \left(a^{\alpha\beta} \frac{\partial p_{\alpha\beta}}{\partial u} + b_\gamma x_u^\gamma \right) + (2a^{13} x_w^3 - a^{33} x_w^1) \frac{\partial W_1}{\partial u} \right\} = 0,$$

which becomes

$$x_w^3 x_u^1 x_u^2 \left\{ (x_w^3)^2 F_u + (2a^{13} x_w^3 - a^{33} x_w^1) \frac{\partial W_1}{\partial u} \right\} = 0, \tag{5.4}$$

when account is taken of (1.7) or (4.16a), and (4.16b).

In a similar manner we obtain from (1.9) the equation

$$x_w^3 \frac{\partial p_{12}}{\partial w} a^{\alpha\beta} D_\alpha D_\beta + x_w^3 x_u^2 \left\{ x_u^1 (x_w^3)^2 \left(a^{\alpha\beta} \frac{\partial p_{\alpha\beta}}{\partial w} + b_\gamma x_w^\gamma \right) - \mathfrak{A}_{22} \frac{\partial W_1}{\partial u} \right\} = 0.$$

When account is taken of (1.7) and $W_2 \equiv W_3 \equiv 0$ this becomes

$$x_w^3 x_u^2 \left\{ \mathfrak{A}_{22} \frac{\partial W_1}{\partial u} - x_u^1 (x_w^3)^2 (F_w - F_{p_1} W_1 - F_z W) \right\} = 0. \tag{5.5}$$

It is easily seen that

$$\frac{\partial W}{\partial u} = x_u^1 W_1 \tag{5.6}$$

when use is made of $W_2 \equiv W_3 \equiv 0$. The system of linear partial differential equations (5.4), (5.5) and (5.6) in the unknowns F, W, W_1 has an identically vanishing solution. In view of condition (a) and (5.3) and the uniqueness theorem of §6 we must have $F \equiv W \equiv W_1 \equiv 0$.

Multiply the first of equations (4.16c) by x_v^2 and the equation (4.16e) by x_u^2 and subtract and we have upon making use of (4.16a) the equation

$$\frac{\mathfrak{A}_{22}}{x_u^1} \frac{\partial p_{11}}{\partial u} + \frac{\mathfrak{A}_{11}}{x_u^2} \frac{\partial p_{12}}{\partial u} = \frac{\mathfrak{A}_{22}}{x_v^1} \frac{\partial p_{11}}{\partial v} + \frac{\mathfrak{A}_{11}}{x_v^2} \frac{\partial p_{12}}{\partial v}.$$

Introducing x^α as the independent variable this becomes

$$\mathfrak{A}_{22} \frac{\partial p_{11}}{\partial x^2} \left(\frac{x_u^2}{x_u^1} - \frac{x_v^2}{x_v^1} \right) = \mathfrak{A}_{11} \frac{\partial p_{12}}{\partial x^1} \left(\frac{x_v^1}{x_v^2} - \frac{x_u^1}{x_u^2} \right).$$

From equations (4.16a) we have

$$\frac{\mathfrak{A}_{22}}{\mathfrak{A}_{11}} = \frac{x_u^1 x_v^1}{x_u^2 x_v^2}$$

and since $x_u^1 x_v^2 - x_u^2 x_v^1 \neq 0$, it follows that

$$\frac{\partial p_{11}}{\partial x^2} = \frac{\partial p_{12}}{\partial x^1}. \tag{5.7}$$

Analogous to (5.1) we have

$$\frac{\partial V_\alpha}{\partial u} = \frac{\partial p_{\alpha\beta}}{\partial v} x^\beta_u - \frac{\partial p_{\alpha\beta}}{\partial u} x^\beta_v,$$

which together with (5.7) implies that $\partial V_1/\partial u = 0$. Thus we have $V_1 \equiv 0$.

We have yet to show the vanishing of V, V_2, V_3. Analogous to (5.6) we have

$$\frac{\partial V}{\partial u} = x^2_u V_2 + x^3_u V_3. \tag{5.8}$$

It is easily verified that $x^\alpha_w \partial V_\alpha/\partial u = x^\alpha_u \partial V_\alpha/\partial w$ when account is taken of $U_\alpha \equiv W_\alpha \equiv 0$. Since $V_1 \equiv 0$ we can then write

$$x^2_w \frac{\partial V_2}{\partial u} + x^3_w \frac{\partial V_3}{\partial u} = x^2_u \frac{\partial V_2}{\partial w} + x^3_u \frac{\partial V_3}{\partial w}. \tag{5.9}$$

Combining equations (1.9) in a third manner we obtain

$$x^3_w \frac{\partial p_{12}}{\partial v} a^{\alpha\beta} D_\alpha D_\beta + x^3_w x^1_u \left\{ (x^3_w)^2 x^2_u \left(a^{\alpha\beta} \frac{\partial p_{\alpha\beta}}{\partial v} + b_\gamma x^\gamma_v \right) - \mathfrak{A}_{11} \frac{\partial V_2}{\partial u} \right.$$
$$\left. - x^2_u \left[(2a^{23} x^3_w - a^{33} x^2_w) \frac{\partial V_2}{\partial w} + a^{33} x^3_w \frac{\partial V_3}{\partial w} \right] \right\} = 0.$$

Making use of (1.7), $F \equiv 0$, and $V_1 \equiv 0$ this becomes

$$x^3_w x^1_u \left\{ \mathfrak{A}_{11} \frac{\partial V_2}{\partial u} + x^2_u \left[(2a^{23} x^3_w - a^{33} x^2_w) \frac{\partial V_2}{\partial w} + x^3_w a^{33} \frac{\partial V_3}{\partial w} \right. \right.$$
$$\left. \left. + (F_z V + F_{p_2} V_2 + F_{p_3} V_3)(x^3_w)^2 \right] \right\} = 0. \tag{5.10}$$

Equations (5.8), (5.9), and (5.10) together with (5.3) imply that $V \equiv V_2 \equiv V_3 \equiv 0$.

6. An existence theorem for linear partial differential equations

Consider a system of n partial differential equations of the form

$$\text{(a)} \quad a^\alpha_i \frac{\partial y_\alpha}{\partial u} = b^\alpha_i \frac{\partial y_\alpha}{\partial w} + c^\alpha_i y_\alpha + d_i, \qquad \text{(b)} \quad a^\alpha_a \frac{\partial y_\alpha}{\partial v} = b^\alpha_a \frac{\partial y_\alpha}{\partial w} + c^\alpha_a y_\alpha + d_a, \qquad \begin{bmatrix} \alpha, \beta, \gamma = 1, \cdots, n \\ i, j = 1, \cdots, p \\ a, b = p+1, \cdots, n \end{bmatrix} \tag{6.1}$$

in which the y_α are to be determined as functions of u, v, w. Here the variables u, v, s, w have the same relation to Euclidean space with rectangular Cartesian coördinates as in §§4, 5 but otherwise the notation used in this section bears no relation to the rest of the paper.

We propose the following initial value problem. We have given over the plane $u + v = 0$, the functions $y^{(0)}_\alpha(s, w)$ which are partially analytic with

respect to w for $w = w_0$ in an interval G^*. The a_β^α, b_β^α, c_β^α, d_α, $\partial a_\beta^\alpha/\partial u$, $\partial a_\beta^\alpha/\partial v$ shall be functions of the three variables u, v, w, partially analytic with respect to w for $w = w_0$ in a two dimensional region G containing G^* in its interior. We can just as well assume that G^* consists of all points of $u + v = 0$ in G. The determinant $| a_\beta^\alpha |$ shall be different from zero in its region of definition. Under these assumptions we shall show that there exists a unique solution $y_\alpha(u, v, w)$ of (6.1), partially analytic with respect to w for $w = w_0$ in a subregion G_1 of G, which takes on the given initial values $y_\alpha^{(0)}(s, w)$ over $u + v = 0$.

As a first step we make the change of unknown

$$z_\beta = a_\beta^\alpha y_\alpha, \tag{6.2}$$

and the system (6.1) takes the form

$$\begin{aligned} \frac{\partial z_i}{\partial u} &= f_i^\alpha \frac{\partial z_\alpha}{\partial w} + g_i^\alpha z_\alpha + d_i, \\ \frac{\partial z_a}{\partial v} &= f_a^\alpha \frac{\partial z_\alpha}{\partial w} + g_a^\alpha z_\alpha + d_a. \end{aligned} \tag{6.3}$$

We adopt the notation

$$z_\alpha^{(0)}(s, w) = a_\alpha^\beta y_\beta^{(0)}(s, w)$$

for the transformed initial values.

We employ a method of successive approximations and define the ν^{th} approximating function z_α^ν by the equations

$$\begin{aligned} z_i^\nu &= z_i^{(0)}(-v, v, w) + \int_{-v}^{u} \left\{ f_i^\alpha \frac{\partial z_\alpha^{\nu-1}}{\partial w} + g_i^\alpha z_\alpha^{\nu-1} + d_i \right\} (u', v, w)\, du', \\ z_a^\nu &= z_a^{(0)}(u, -u, w) + \int_{-u}^{v} \left\{ f_a^\alpha \frac{\partial z_\alpha^{\nu-1}}{\partial w} + g_a^\alpha z_\alpha^{\nu-1} + d_a \right\} (u, v', w)\, dv'. \end{aligned} \tag{6.4}$$

By $z_i^{(0)}(-v, v, w)$ we mean the assigned initial value $z_i^{(0)}(s, w)$ at the point on the plane $u + v = 0$ with coördinates $-v, v, w$. The variables u' and v' which appear as arguments in the integrand are variables of integration. To complete our definition of z_α^ν we take

$$\begin{aligned} z_i^0 &= z_i^{(0)}(-v, v, w), \\ z_a^0 &= z_a^{(0)}(u, -u, w). \end{aligned} \tag{6.5}$$

The $z_\alpha^\nu(u, v, w)$ defined by (6.4) and (6.5) will be partially analytic with respect to w for $w = w_0$ in the region G. In order to be able to perform the integrations indicated in (6.4) we must restrict G to be of such a shape that if u, v lies in G so does the entire triangle with vertices (u, v), $(u, -u)$, $(-v, v)$.

For convenience we take $w_0 = 0$ and choose positive constants m and r so that

$$\frac{m}{\left(1 - \frac{w}{r}\right)} \tag{6.6}$$

dominates each of the coefficients f_β^α, g_β^α, d_α and $z_\alpha^{(0)}$ for each pair of values u, v in G. Let $f(w)$ be any function analytic in the neighborhood of $w = 0$ such that

$$f(w) \ll \frac{M}{\left(1 - \frac{w}{r}\right)^h},$$

where $\ll$ means "dominated by." Then it follows that

$$f'(w) \ll \frac{Mh}{r\left(1 - \frac{w}{r}\right)^{h+1}}. \tag{6.7}$$

Let us note that

$$\overset{\nu}{z}_\alpha = \overset{0}{z}_\alpha + \sum_{\mu=1}^{\nu} (\overset{\mu}{z}_\alpha - \overset{\mu-1}{z}_\alpha). \tag{6.8}$$

From (6.4), (6.6), and (6.7) we have

$$\overset{1}{z}_\alpha - \overset{0}{z}_\alpha \ll \frac{M_1}{\left(1 - \frac{w}{r}\right)^3} \,|\, u + v \,| \qquad \text{where } M_1 = m\left(1 + nm + \frac{nm}{r}\right).$$

In turn we get

$$\overset{\mu}{z}_\alpha - \overset{\mu-1}{z}_\alpha \ll \frac{M_\mu}{\left(1 - \frac{w}{r}\right)^{2\mu+1}} \,|\, u + v \,|^\mu \qquad (\mu = 1, 2, \cdots), \tag{6.9}$$

where

$$M_\mu = \frac{nmM_{\mu-1}}{\mu}\left(\frac{2\mu - 1}{r} + 1\right) \qquad (\mu = 2, 3, \cdots).$$

Put

$$\operatorname*{Limit}_{\mu \to \infty}\left(\frac{M_\mu}{M_{\mu-1}}\right) = \frac{2nm}{r} = m'.$$

Select a positive constant $k < r$ and let G_1 denote a subregion of G which satisfies the inequality

$$\frac{m' \,|\, u + v \,|}{\left(1 - \frac{k}{r}\right)^2} \leqq \delta < 1. \tag{6.10}$$

Since $|u + v|$ is proportional to the distance from (u, v, w) to the plane $u + v = 0$ this inequality requires that we remain sufficiently close to the initial plane. In view of (6.9) each term of (6.8) will be in absolute value less than the corresponding term of the convergent series of positive constants

$$\frac{1}{\left(1 - \frac{k}{r}\right)}\left[m + \sum_{\mu=1}^{\infty} M_\mu \left(\frac{\delta}{m'}\right)^\mu\right]$$

for values of u, v, in G_1 and values of w satisfying $|w| \leqq k$.

Hence, the function $z_\alpha^\nu(u, v, w)$ converges uniformly to a continuous limit function $z_\alpha(u, v, w)$. Furthermore making use of the theorem on absolute convergence of double arrays the limit function $z_\alpha(u, v, w)$ will be partially analytic with respect to w for $w = w_0$ in the region G_1. Let $\nu \to \infty$ in (6.4) we obtain

$$\text{(6.11)}\quad \begin{aligned} &\text{(a)}\quad z_i = z_i^{(0)}(-v, v, w) + \int_{-v}^{u} \left\{ f_i^\alpha \frac{\partial z_\alpha}{\partial w} + g_i^\alpha z_\alpha + d_i \right\} du', \\ &\text{(b)}\quad z_a = z_a^{(0)}(u, -u, w) + \int_{-u}^{v} \left\{ f_a^\alpha \frac{\partial z_\alpha}{\partial w} + g_a^\alpha z_\alpha + d_a \right\} dv'. \end{aligned}$$

The functions $z_\alpha(u, v, w)$ satisfy the initial conditions. Differentiating (6.11a) with respect to u and (6.11b) with respect to v we see that the system (6.3) is satisfied.

Consider the uniqueness of the solution. Let $z_\alpha^{(1)}$ and $z_\alpha^{(2)}$ be two such solutions of the system (6.3) which take on the same initial values $z_\alpha^{(0)}$ over the plane $u + v = 0$. We will show that in a neighborhood of $u + v = 0$, $z_\alpha^{(1)} \equiv z_\alpha^{(2)}$. The functions $z_\alpha^{(1)}$ and $z_\alpha^{(2)}$ must both satisfy the system (6.11). Select the constants m and r so that (6.6) dominates the coefficients of (6.3) and also the two solutions $z_\alpha^{(1)}$ and $z_\alpha^{(2)}$. Replace the z_α in (6.11) by $z_\alpha^{(1)}$ and $z_\alpha^{(2)}$, subtract corresponding equations and obtain in succession

$$z_\alpha^{(1)} - z_\alpha^{(2)} \ll \frac{M_\mu}{\left(1 - \frac{w}{r}\right)^{2\mu+1}} |u + v|^\mu \qquad (\mu = 1, 2, \ldots),$$

where

$$M_1 = 2nm^2\left(1 + \frac{1}{r}\right), \qquad M_\mu = \frac{nmM_{\mu-1}}{\mu}\left(\frac{2\mu - 1}{r} + 1\right) \qquad (\mu = 2, 3, \cdots).$$

Select G_1 as before and obtain

$$|z_\alpha^{(1)} - z_\alpha^{(2)}| < \frac{M_\mu}{\left(1 - \frac{k}{r}\right)}\left(\frac{\delta}{m'}\right)^\mu \qquad (\mu = 1, 2, \cdots).$$

In the limit we obtain $z_\alpha^{(1)} \equiv z_\alpha^{(2)}$ for u and v in G_1 and $|w| \leqq k$.

7. Initial value problem for a linear equation $\bar{F} = 0$

In case (1.1) is the linear differential equation

$$F \equiv a^{\alpha\beta} p_{\alpha\beta} + c^{\gamma} p_{\gamma} + dz + e = 0, \tag{7.1}$$

where the coefficients $a^{\alpha\beta}$, c^{γ}, d, e are functions of x^{α} alone, the analysis of §6 can be used to show both the existence and in a sense the uniqueness of a solution of the initial value problem for $\bar{F} = 0$.

In the case of the linear equation (7.1) the foregoing theory can be considerably simplified. The $p_{\alpha\beta}$ can be dropped from the strip (1.2) and the corresponding strip conditions (1.3) can be omitted. Then the differentiation (1.4) is unnecessary and the last column of (1.6) is replaced by b, $-\partial p_1/\partial u$, $-\partial p_1/\partial w$, $-\partial p_2/\partial u$, $-\partial p_2/\partial w$, $-\partial p_3/\partial u$, $-\partial p_3/\partial w$, where $b = c^{\gamma} p_{\gamma} + dz + e$. The equations (1.9) and (1.10) are correspondingly reduced from six to two equations. The characteristic surfaces of (7.1) are determined independently of a solution z^* and let us examine §4. Since the $a^{\alpha\beta}$ are analytic in their arguments x^{α}, the functions in (4.3) can also be taken to be analytic in their arguments. When one inspects the changes of variables used in obtaining (4.13) it is seen that the functions x^{α} in (4.13) are partially analytic with respect to w. Since the mapping (4.13) can be obtained independently of a solution z^* of $\bar{F} = 0$ we exclude the equations (4.16a) from the system Σ. The system Σ is replaced by the system Σ' containing four equations

$$\begin{aligned}
& z_u = p_{\alpha} \overset{\alpha}{x}_u, \\
& \overset{3}{x}_w \left(\mathfrak{A}_{22}\, \rho \frac{\partial p_1}{\partial u} + \mathfrak{A}_{11} \frac{\partial p_2}{\partial u} \right) + \overset{2}{x}_u E = 0, \\
& \overset{\alpha}{x}_w \frac{\partial p_{\alpha}}{\partial u} - \overset{\alpha}{x}_u \frac{\partial p_{\alpha}}{\partial w} = 0, \\
& \overset{3}{x}_w \left(\mathfrak{A}_{22}\, \sigma \frac{\partial p_1}{\partial v} + \mathfrak{A}_{11} \frac{\partial p_2}{\partial v} \right) + \overset{2}{x}_v E = 0,
\end{aligned} \tag{Σ'}$$

where E is linear in $\partial p_{\alpha}/\partial w$, p_{α}, z. Remembering that the x^{α} are known functions of u, v, w it is seen that Σ' is linear in the z, p_{α}, and their derivatives with respect to u, v, w with coefficients partially analytic with respect to w. In fact the system Σ' is of the type considered in §6 when we note that the determinant

$$\begin{vmatrix} \rho\, \mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 \\ \overset{1}{x}_w & \overset{2}{x}_w & \overset{3}{x}_w \\ \sigma\, \mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 \end{vmatrix} = \overset{3}{x}_w (\mathfrak{A}_{22} \mathfrak{A}_{11})^2 (\sigma - \rho). \tag{7.2}$$

The non-vanishing of (7.2) over $u + v = 0$ follows immediately from condition (c′) since ρ and σ must be different and both different from zero.

By the existence theorem in §6 there exists a unique solution of the initial value problem for Σ', which is partially analytic with respect to w. By an argument[13] similar to that in §5 it can be shown that this solution of Σ' is also a solution of the initial value problem for $\bar{F} = 0$. Conversely, consider a solution of the initial value problem for $\bar{F} = 0$ which furnishes a solution of Σ' partially analytic with respect to w when the substitution (4.13) has been made. By the uniqueness theorem of §6 this solution of Σ' is unique.

8. An existence theorem for non-linear partial differential equations

Consider a system of n partial differential equations of the form

$$\text{(8.1)}\quad \begin{aligned} &\text{(a)}\quad \Phi_i \equiv a_i^\alpha \frac{\partial y_\alpha}{\partial u} - b_i = 0,\\ &\text{(b)}\quad \Phi_a \equiv a_a^\alpha \frac{\partial y_\alpha}{\partial v} - b_a = 0, \end{aligned} \qquad \begin{bmatrix} \alpha, \beta, \gamma, = 1, \cdots, n \\ i, j = 1, \cdots, p \\ a, b = p + 1, \cdots, n \end{bmatrix}$$

in which the y_α are to be determined as functions of u, v, w. We have given over the plane $u + v = 0$ the initial values $y_\alpha^{(0)}(s, w)$. The functions $y_\alpha^{(0)}(s, w)$ and their first derivatives with respect to s shall be partially analytic with respect to w for $w = w_0$ in an interval G^*. The coefficients a_β^α and b_α shall be analytic functions of their arguments y_α and $\partial y_\alpha/\partial w$ in a neighborhood of the set of initial values $y_\alpha^{(0)}$ and $\partial y_\alpha^{(0)}/\partial w$. The determinant $| a_\beta^\alpha |$ shall be different from zero for the given initial values. We shall show that there exists a unique solution $y_\alpha(u, v, w)$ of (8.1) which together with its u, v, and uv derivatives is partially analytic with respect to w for $w = w_0$ in a region G_2 and which takes on the given initial values over $u + v = 0$.

First we differentiate (8.1a) with respect to v and (8.1b) with respect to u, regarding the y_α as functions of u, v, w, and solve the resulting system to obtain

$$\text{(8.2)}\qquad \frac{\partial^2 y_\alpha}{\partial u \partial v} = f_\alpha\left(\frac{\partial^2 y_\alpha}{\partial u \partial w}, \frac{\partial^2 y_\alpha}{\partial v \partial w}, \frac{\partial y_\alpha}{\partial u}, \frac{\partial y_\alpha}{\partial v}, \frac{\partial y_\alpha}{\partial w}, y_\alpha\right).$$

Before integrating the equations (8.2) we shall need to know the initial values $\partial y_\alpha^{(0)}/\partial u$ and $\partial y_\alpha^{(0)}/\partial v$. We have

$$\text{(8.3)}\qquad \sqrt{2}\, \frac{\partial y_\alpha^{(0)}}{\partial s} = \frac{\partial y_\alpha^{(0)}}{\partial u} - \frac{\partial y_\alpha^{(0)}}{\partial v},$$

---

[13] It is to be noted that the argument of §5 requires that the solution of Σ' have continuous u, v, and uv derivatives. The existence of these derivatives could have been insured by considering the system obtained from (6.1) by differentiation with respect to u and v as is done in §8. However, we thought it not advisable to include this complication in our first existence theorem.

It might also be noted at this point that each of the systems in §5 to which the uniqueness theorem of §6 is applied is of the type (6.1a) with (6.1b) vacuous. For such a system the transformation (6.2) is unnecessary and we need not require that the coefficients a_β^α in (6.1) have u and v derivatives.

since $s = \sqrt{2}\, u = -\sqrt{2}\, v$. In addition the initial values $\partial y_\alpha^{(0)}/\partial u$ and $\partial y_\alpha^{(0)}/\partial v$ must satisfy (8.1) and making use of $| a_\beta^\alpha | \neq 0$ we see that the determinant of the coefficients of these $2\, n$ equations, which is of the form

$$\left|\begin{array}{ccc|ccc} & a_i^\alpha & & & 0 & \\ \hline & 0 & & & a_\alpha^\alpha & \\ \hline 1 & & & -1 & & \\ & 1 & 0 & & -1 & 0 \\ & & \ddots & & & \ddots \\ 0 & & 1 & 0 & & -1 \end{array}\right|,$$

is different from zero. We designate the initial values thus obtained by $y_{\alpha u}^{(0)}(s, w)$ and $y_{\alpha v}^{(0)}(s, w)$.

Again we employ a method of successive approximations

$$\text{(8.4)} \quad \begin{aligned} &\text{(a)} \quad y_{\alpha v}^{\nu+1} = y_{\alpha v}^{(0)}(-v, v, w) + \int_{-v}^{u} f_\alpha(y_{uw}^\nu, y_{vw}^\nu, y_u^\nu, y_v^\nu, y_w^\nu, y^\nu)\, du', \\ &\text{(b)} \quad y_{\alpha u}^{\nu+1} = y_{\alpha u}^{(0)}(u, -u, w) + \int_{-u}^{v} f_\alpha(y_{uw}^\nu, y_{vw}^\nu, y_u^\nu, y_v^\nu, y_w^\nu, y^\nu)\, dv', \end{aligned}$$

where for convenience the subscript α is omitted in the integrand but otherwise the notation follows §6. Out of these $2n$ equations we can calculate the functions $y_{\alpha u}^{\nu+1}$ and $y_{\alpha v}^{\nu+1}$ providing the corresponding functions of the superscript ν remain within the region of analyticity of f_α. We first show that from (8.4) we get a unique function $y_\alpha^{\nu+1}$. If

$$\text{(8.5)} \quad \begin{aligned} y_\alpha^{\nu+1} &= y_\alpha^{(0)}(-v, v, w) + \int_{-v}^{u} y_{\alpha u}^{\nu+1}(u', v, w)\, du', \\ \bar{y}_\alpha^{\nu+1} &= y_\alpha^{(0)}(u, -u, w) + \int_{-u}^{v} y_{\alpha v}^{\nu+1}(u, v', w)\, dv', \end{aligned}$$

we shall show that $y_\alpha^{\nu+1} = \bar{y}_\alpha^{\nu+1}$. Substituting from (8.4) we have

$$\text{(8.6)} \quad \begin{aligned} y_\alpha^{\nu+1} = y_\alpha^{(0)}(-v, v, w) &+ \int_{-v}^{u} y_{\alpha u}^{(0)}(u', -u', w)\, du' \\ &+ \int_{-v}^{u} \left\{ \int_{-u'}^{v} f_\alpha(y_{uw}^\nu, \cdots, y^\nu)\, dv' \right\} du'. \end{aligned}$$

But from (8.3) we have

$$\begin{aligned} y_\alpha^{(0)}(-v, v, w) + \int_{-v}^{u} y_{\alpha u}^{(0)}(u', -u', w)\, du' &= y_\alpha^{(0)}(-v, v, w) + \int_{-v}^{u} \Big\{ y_{\alpha v}^{(0)}(u', -u', w) \\ + \sqrt{2}\, y_{\alpha s}^{(0)}(u', -u', w) \Big\}\, du' &= y_\alpha^{(0)}(u, -u, w) + \int_{-u}^{v} y_{\alpha v}^{(0)}(-v', v', w)\, dv', \end{aligned}$$

where the last integral is the result of a change of variable $v' = -u'$. The repeated integral in (8.6) and

$$\int_u^v \left\{ \int_{-v'}^u f_\alpha(y^\nu_{uw}, \cdots, y^\nu)\, du' \right\} dv'$$

are each equal to the integral over the triangle with vertices $(u, -u, w)$, $(-v, v, w)$, and (u, v, w). Hence $y_\alpha^{\nu+1} = \bar{y}_\alpha^{\nu+1}$. We fix the zeroth approximating functions by

(8.7) $y_\alpha^0 = y_\alpha^{(0)}(-v, v, w); \qquad y_{\alpha u}^0 = y_{\alpha u}^{(0)}(u, -u, w); \qquad y_{\alpha v}^0 = y_{\alpha v}^{(0)}(-v, v, w).$

We employ the method of dominant functions in order to show that the process (8.4) and (8.5) can be continued indefinitely. First for convenience we take $w_0 = 0$ in our hypothesis. By limiting ourselves to values of u and v for which $|u + v|$ is sufficiently small, the function $y_\alpha^1(u, v, w)$ and its derivatives will remain within the region of analyticity of f_α. Furthermore by (8.4) and (8.5) the function $y_\alpha^1(u, v, w)$ reduces to $y_\alpha^{(0)}(s, w)$ for $u + v = 0$ and its derivatives $y_{\alpha u}^1(u, v, w)$ and $y_{\alpha v}^1(u, v, w)$ to $y_{\alpha u}^{(0)}(s, w)$ and $y_{\alpha v}^{(0)}(s, w)$ for $u + v = 0$. Hence if we make the change of unknown

(8.8) $$z_\alpha = y_\alpha - y_\alpha^1(u, v, w),$$

we obtain in place of (8.2) the equations

(8.9) $$\frac{\partial^2 z_\alpha}{\partial u \partial v} = F_\alpha(z_{uw}, z_{vw}, z_u, z_v, z_w, z, u, v, w)$$

with the initial conditions $z_\alpha^{(0)}(s, w) \equiv 0$, $z_{\alpha u}^{(0)}(s, w) \equiv 0$, $z_{\alpha v}^{(0)}(s, w) \equiv 0$. The function F_α is partially analytic with respect to all its arguments (except u and v) for the zero values of these arguments in a region G of the space of coördinates u and v. Let us select positive constants M and R so that for values of u and v in G the right member of

(8.10) $$\frac{\partial^2 Z_\alpha}{\partial u \partial v} = \frac{M}{1 - \dfrac{n(Z_{uw} + Z_{vw} + Z_u + Z_v + Z_w + Z) + w}{R}}$$

will dominate the right member of (8.9). Take $\sigma = u + v$ and look for Z as a function of σ and w; then (8.10) becomes

(8.11) $$\frac{\partial^2 Z_\alpha}{\partial \sigma^2} = \frac{M}{1 - \dfrac{n(2Z_{\sigma w} + 2Z_\sigma + Z_w + Z) + w}{R}}$$

with initial conditions $Z_\alpha = 0$ and $\partial Z_\alpha / \partial \sigma = 0$ for $\sigma = 0$. By the Cauchy Kowalewski theorem (8.11) has a solution $Z_\alpha(\sigma, w)$ which has a convergent power series expansion about $(0, 0)$ and which together with its derivative $\partial Z_\alpha / \partial \sigma$ reduces to zero for $\sigma = 0$. It is evident that the coefficients of this

power series expansion are positive. Hence, when we replace σ by $u + v$ we obtain a solution $Z_\alpha(u + v, w)$ of (8.10) which has a power series expansion in $u + v$ and w with positive coefficients convergent, say for $|u + v| \leqq \gamma$, $|w| \leqq \delta$ and which together with its u and v derivatives reduces to zero for $u + v = 0$. Let us choose a subregion G_1 of G which satisfies $|u + v| \leqq \gamma$. Also, the Z_α and their derivatives satisfy the integral equations

$$\text{(a)} \quad Z_{\alpha v} = \int_{-v}^{u} \frac{M}{1 - \dfrac{n(Z_{uw} + \cdots + Z) + w}{R}}\, du',$$

$$\text{(8.12)} \qquad \text{(b)} \quad Z_{\alpha u} = \int_{-u}^{v} \frac{M}{1 - \dfrac{n(Z_{uw} + \cdots + Z) + w}{R}}\, dv',$$

$$\text{(c)} \quad Z_\alpha = \int_{-v}^{u} Z_{\alpha u}(u', v, w)\, du'.$$

Let us suppose that the method of successive approximations has been carried out on the system (8.9) and we shall refer to the equations corresponding to (8.4), (8.5), etc., as (8.4*), (8.5*), etc. We now assume that

(α) Each of the functions $z_\alpha^\nu(u, v, w)$; $z_{\alpha u}^\nu(u, v, w)$; $z_{\alpha v}^\nu(u, v, w)$ is partially analytic with respect to w for $w = 0$ for values of u and v in G_1.

(β) The relations

$$Z_\alpha(|u + v|, w) \gg z_\alpha^\nu(u, v, w),$$
$$Z_{\alpha u}(|u + v|, w) \gg z_{\alpha u}^\nu(u, v, w),$$
$$Z_{\alpha v}(|u + v|, w) \gg z_{\alpha v}^\nu(u, v, w),$$

are valid in G_1.

We shall show (α) and (β) are valid with ν replaced by $\nu + 1$. Let us use

$$\text{(8.13)} \qquad \frac{M}{1 - \dfrac{n(Z_{uw} + \cdots + Z) + w}{R}} = \sum_\mu A_\mu(|u + v|)\, w^\mu$$

to designate the result of replacing the Z_α and their derivatives in the power series expansion for the right member of (8.10) by the expansions for $Z_\alpha(|u + v|, w)$, etc., in powers of w. Similarly let us use

$$\text{(8.14)} \qquad F_\alpha(z_{uw}^\nu, \cdots, z^\nu, u, v, w) = \sum_\mu B_{\alpha\mu}(u, v) w^\mu$$

to designate the result of replacing the $z_\alpha^\nu, \cdots, z_{\alpha uw}^\nu$, in the expansion for F_α, by their expansions in powers of w. Due to property (β) and the fact that the right member of (8.10) dominates the right member of (8.9) the series in the right member of (8.14) will converge for u and v in G_1. In fact we can write

$$\text{(8.15)} \qquad \sum_\mu A_\mu(|u + v|)\, w^\mu \gg \sum_\mu B_{\alpha\mu}(u, v)\, w^\mu$$

for u and v in G_1. Put $A_\mu(|u + v|) = \sum_\lambda C_{\mu\lambda} |u + v|^\lambda$ where $C_{\mu\lambda} \geqq 0$. From (8.15) we get

$$\left| \int_{-u}^{v} B_{\alpha\mu}(u, v')\, dv' \right| \leqq \left| \int_{-u}^{v} A_\mu(|u + v'|)\, dv' \right| = \sum_\lambda C_{\lambda\mu} \frac{|u + v|^{\lambda+1}}{\lambda + 1}. \tag{8.16}$$

From (8.12b) it follows that the extreme right member of (8.16) is the coefficient of w^μ in the expansion for $Z_{\alpha u}(|u + v|, w)$. From (8.4b*) it follows that the quantity within the absolute sign in the extreme left member of (8.16) is the coefficient of w^μ in the expansion of $z_{\alpha u}^{\nu+1}$. In view of the above argument it is easily seen that properties (α) and (β) hold for ν replaced by $\nu + 1$ so far as the function $z_{\alpha u}^\nu$ is concerned. Similar arguments can be given for the functions z_α^ν and $z_{\alpha v}^\nu$. From the fact that (8.7*) are identically zero and the fact that any power series with positive coefficients dominates the function zero, it follows that (α) and (β) hold for $\nu = 0$. Hence, the process (8.4*) and (8.5*) can be continued indefinitely.

We have yet to show that the functions z_α^ν, $z_{\alpha u}^\nu$, $z_{\alpha v}^\nu$ tend uniformly to continuous limit functions. Employing the mean value theorem we get from (8.4*) and (8.5*)

$$\begin{aligned} z_{\alpha v}^{\mu+1} - z_{\alpha v}^{\mu} &= \int_{-v}^{u} \left\{ \frac{\partial F_\alpha}{\partial z_{\beta uw}} (z_{\beta uw}^{\mu} - z_{\beta uw}^{\mu-1}) + \cdots + \frac{\partial F_\alpha}{\partial z_\beta} (z_\beta^\mu - z_\beta^{\mu-1}) \right\} du', \\ z_{\alpha u}^{\mu+1} - z_{\alpha u}^{\mu} &= \int_{-u}^{v} \left\{ \frac{\partial F_\alpha}{\partial z_{\beta uw}} (z_{\beta uw}^{\mu} - z_{\beta uw}^{\mu-1}) + \cdots + \frac{\partial F_\alpha}{\partial z_\beta} (z_\beta^\mu - z_\beta^{\mu-1}) \right\} dv', \\ z_\alpha^{\mu+1} - z_\alpha^\mu &= \int_{-v}^{u} (z_{\alpha u}^{\mu+1} - z_{\alpha u}^{\mu})\, du'. \end{aligned} \tag{8.17}$$

If we let ζ_σ denote the set of arguments $z_{\alpha uw}, \cdots, z_\alpha$ in the function F_α then each of the coefficients $\partial F/\partial \zeta_\sigma$ in (8.17) is evaluated at $\theta_1 \zeta_\sigma^\mu + \theta_2 \zeta_\sigma^{\mu-1}$ where θ_1 and θ_2 are positive constants satisfying $\theta_1 + \theta_2 = 1$. Since the functions $z_{\alpha uw}^\mu, \cdots, z_\alpha^\mu$ are dominated by $Z_{\alpha uw}(|u + v|, w), \cdots, Z_\alpha(|u + v|, w)$ respectively for values of u and v in G_1, we can obtain one function

$$\frac{m}{1 - \dfrac{w}{r}}, \tag{8.18}$$

which dominates $F_\alpha(\zeta_\sigma^0 \mid u, v, w)$, $\dfrac{\partial F_\alpha}{\partial \zeta_\sigma} (\theta_1 \zeta_\sigma^\mu + \theta_2 \zeta_\sigma^{\mu-1} \mid u, v, w)$ for the same values of u and v; m and r are positive constants. From (8.4*) and (8.18) we get

$$z_{\alpha v}^1 \ll \frac{m}{1 - \dfrac{w}{r}} |u + v|,$$

$$z_{\alpha u}^1 \ll \frac{m}{1 - \dfrac{w}{r}} |u + v|,$$

and in turn (8.5*) yields

$$z_\alpha^1 \ll \frac{m}{1 - \frac{w}{r}} \frac{|u + v|^2}{2}. \tag{8.19}$$

For convenience in calculation we suppose that in addition to the conditions so far imposed on $|u + v|$ that $|u + v| < 1$ and (8.19) can be replaced by

$$z_\alpha^1 \ll \frac{m}{1 - \frac{w}{r}} |u + v|.$$

In a similar manner we obtain from (8.17)

$$(z_{\alpha v}^\mu - z_{\alpha v}^{\mu-1}) \ll \frac{M_\mu}{\left(1 - \frac{w}{r}\right)^{2\mu-1}} |u + v|^\mu,$$

$$(z_{\alpha u}^\mu - z_{\alpha u}^{\mu-1}) \ll \frac{M_\mu}{\left(1 - \frac{w}{r}\right)^{2\mu-1}} |u + v|^\mu, \qquad (\mu = 1, 2, 3, \cdots)$$

$$(z_\alpha^\mu - z_\alpha^{\mu-1}) \ll \frac{M_\mu}{\left(1 - \frac{w}{r}\right)^{2\mu-1}} |u + v|^\mu,$$

where

$$M_\mu = \frac{3nmM_{\mu-1}}{\mu}\left(\frac{2\mu - 3}{r} + 1\right), \qquad \begin{array}{l} M_1 = m, \\ (\mu = 2, 3, \cdots). \end{array}$$

If we let $m' = 6nm/r$ and let G_2 denote a subregion of G_1 which satisfies in addition to (6.10) the inequality $|u + v| < 1$, we obtain limit functions z_α, $z_{\alpha u}$, $z_{\alpha v}$, partially analytic with respect to w for $w = 0$ for values of u and v in G_2, which satisfy the equations

$$\begin{aligned} z_{\alpha u} &= \int_{-u}^{v} F_\alpha(z_{uw}, \cdots, z, u, v, w)\, dv', \\ z_{\alpha v} &= \int_{-v}^{u} F_\alpha(z_{uw}, \cdots, z, u, v, w)\, du', \\ z_\alpha &= \int_{-v}^{u} z_{\alpha u}\, du'. \end{aligned} \tag{8.20}$$

Differentiating (8.20) we see that the functions $z_\alpha(u, v, w)$ possess continuous derivatives $z_{\alpha uv}$ partially analytic with respect to w for $w = 0$ for u and v in G_2. The uniqueness of the solution of the system (8.20) is established in the usual manner.

The transformation (8.8) yields the set of functions $y_\alpha(u, v, w)$ which we set

out to find. For these functions $y_\alpha(u, v, w)$ satisfy (8.2) and the initial conditions. If we let $\Phi_\alpha(u, v, w)$ denote the result of substituting the functions $y_\alpha(u, v, w)$ into the left members of (8.1), it follows from (8.2) that

$$\frac{\partial \Phi_i}{\partial v} = 0, \qquad \frac{\partial \Phi_a}{\partial u} = 0. \tag{8.21}$$

The system (8.21) with initial conditions $\Phi_\alpha = 0$ for $u + v = 0$ has by the uniqueness theorem in §6 the unique solution $\Phi_\alpha \equiv 0$.

9. Initial value problem for a non-linear equation $\bar{F} = 0$

The analysis of §8 can be used to show the existence and in a sense the uniqueness of a solution for the initial value problem for $\bar{F} = 0$.

We note that the system Σ is of the type considered in §8 since the determinant

$$\begin{vmatrix} \rho\mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 & 0 & 0 & 0 \\ 0 & \rho\mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho\mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 \\ 0 & x_w^1 & x_w^2 & 0 & x_w^3 & 0 \\ 0 & 0 & 0 & x_w^1 & x_w^2 & x_w^3 \\ \sigma\mathfrak{A}_{22} & \mathfrak{A}_{11} & 0 & 0 & 0 & 0 \end{vmatrix} = (\mathfrak{A}_{11}\mathfrak{A}_{22}x_w^3)^2(\rho^2 - \rho\sigma) \neq 0 \tag{9.1}$$

over $u + v = 0$. By the existence theorem in §8 there exists a unique solution of the initial value problem for Σ which is partially analytic with respect to w. By the argument in §5 this solution of Σ is also a solution of the initial value problem for $\bar{F} = 0$. Conversely, any solution of the initial value problem for $\bar{F} = 0$ which admits a mapping (4.14) partially analytic with respect to w is unique.

UNIVERSITY OF MARYLAND.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

THE RIEMANNIAN AND AFFINE DIFFERENTIAL GEOMETRY OF PRODUCT-SPACES

By F. A. Ficken

(Received August 19, 1938)

A Riemannian geometry is completely determined by defining over a space a quadratic differential form $ds^2 = g_{ab}dx^a dx^b$, called the metric form. Let $\{P\}$ and $\{Q\}$ denote two Riemannian spaces, of dimensions p and q, with metric forms whose coefficients are g_{ab} and g_{ij}. Then the product $\{P\} \times \{Q\}$ is a well-defined space $\{R\}$ of dimension $r = p + q$. A metric may be assigned to $\{R\}$ at will, but, in order that the geometry of $\{R\}$ may be accessible through the geometries of $\{P\}$ and $\{Q\}$, a metric is suggested which depends on the given metrics of $\{P\}$ and $\{Q\}$. If the metric of $\{R\}$ has coefficients

$$g_{\alpha\beta} = \begin{pmatrix} g_{ab} & 0 \\ 0 & g_{ij} \end{pmatrix},$$

then the geometry produced on $\{R\}$ may be inferred in great detail from the given geometries of $\{P\}$ and $\{Q\}$. To examine this inference is the main purpose of the present paper.

A geometric object S in a product-space may be called a product-object (§2) if, symbolically,

$$S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix},$$

where S_1 and S_2 are similar objects on the factor spaces. Thus $g_{\alpha\beta}$ is defined above to be a product-tensor, and $\Gamma^\alpha_{\beta\gamma}$, $R^\alpha_{\beta\gamma\delta}$, etc. are product-objects. A product-space has families of fundamental subspaces, each of which is a copy of one of the factor-spaces. Any subspace of a product-space has projections (§4) on these fundamental subspaces. In §§4–7 product-objects and projections are used to exhibit the relation between the geometry of our product space and that of its factors with regard to geodesic subspaces, parallel displacement, curvature, parallel fields of vector-spaces, and motions. These results are extended in §8 to the product of two affinely connected spaces. An indication is given in §9 of the type of difficulty which arises in connection with other definitions of $g_{\alpha\beta}$.

It is an unsolved problem to give useful necessary and sufficient conditions that a Riemann space be a product-space. Little is known of product-spaces in the large, or of products of spaces with an indefinite metric. None of these problems is treated in this paper.

All functions are assumed to have differentiability properties adequate to the part they play in the discussion. In a general way, notation and terminology follow L. P. Eisenhart's texts.[1]

The author is indebted to Professor A. W. Tucker for a suggestion which led to the study of this problem. Professors L. P. Eisenhart and M. S. Knebelman have kindly examined the manuscript, and have offered many very valuable suggestions. For this generous and stimulating counsel the author eagerly extends his warm thanks.

I. Introductory Remarks

1. Let $\{P\}$ and $\{Q\}$ be two topological spaces. We form the set $\{R\}$ of ordered pairs $R = (P, Q)$, where P is a point in $\{P\}$ and Q a point in $\{Q\}$, and agree that the point R shall be "near" the point R' if and only if P is "near" P' and Q "near" Q'. Then $\{R\}$ is a topological space which is called the product of $\{P\}$ and $\{Q\}$; we write $\{R\} = \{P\} \times \{Q\}$. Examples: With any of the usual meanings of "nearness," the product of two straight lines is a plane, the product of two circles is a torus, and the product of a 2-sphere and a line segment is a spherical shell.

If $\{P\}$ and $\{Q\}$ are manifolds of dimensions p and q and classes u and v, then $\{R\}$ is a manifold of dimension $r = p + q$ and of class w equal to the lesser of u and v.

Except in §8, the present discussion is conducted on the hypothesis that $\{P\}$ and $\{Q\}$ are Riemann spaces $\mathfrak{R}_p$ and $\mathfrak{R}_q$. Indices are assigned the following permanent ranges: $\alpha = 1, \cdots, r$; α^1, $a = 1, \cdots, p$; and α^2, $i = 1, \cdots, q$ or $p+1, \cdots, r$ according to context. α^1 and a are indices of the *first kind*, α^2 and i of the *second*. It is usually convenient to use Greek indices in $\{R\}$ and Latin indices in $\mathfrak{R}_p$ and $\mathfrak{R}_q$.

If $P \leftrightarrow x^a$ and $Q \leftrightarrow x^i$ are coördinate systems for $\mathfrak{R}_p$ and $\mathfrak{R}_q$, then $R \leftrightarrow x^\alpha$ is a coördinate system for $\{R\}$ which will be described as the product of the two given coördinate systems. Such a coördinate system will be called a code, and codes will be used exclusively. Every transformation of coördinates (change of code) in $\{R\}$ is understood, therefore, to be of the form

$$\bar{x}^{\alpha^1} = f^{\alpha^1}(x^1, \cdots, x^p)$$

$$\bar{x}^{\alpha^2} = f^{\alpha^2}(x^{p+1}, \cdots, x^r).$$

Equations $x^{\alpha^1} = c^{\alpha^1}$ define subspaces of $\{R\}$ which are topologically equivalent to the second factor-space $\mathfrak{R}_q$; any such subspace will be denoted by R_q. The equations $\bar{x}^{\alpha^1} = x^{\alpha^1} + t^{\alpha^1}$ set up a regular one-one correspondence of class v between the R_q given by $x^{\alpha^1} = c^{\alpha^1}$ and the R_q given by $x^{\alpha^1} = c^{\alpha^1} + t^{\alpha^1}$. The equations $x^{\alpha^2} = c^{\alpha^2}$ give an analogous system of subspaces R_p.

A manifold such as $\{R\}$ is (locally) the topological product of factor-manifolds

[1] *Riemannian Geometry*, cited as R.G., and *Continuous Groups of Transformations*, cited as C.G.

in several different ways for each coördinate system (not necessarily a code, of course). For the present discussion, it is desired to define a positive definite metric tensor $g_{\alpha\beta}$ of class $w - 1$ over $\{R\}$ in such a way that the resulting Riemann space R_r may justly be called a metric product of $\mathfrak{R}_p$ and $\mathfrak{R}_q$.

It seems natural to require that the geometry assigned by $g_{\alpha\beta}$ to R_r shall induce on each subspace R_p (and R_q) the geometry already given on $\mathfrak{R}_p$ (and $\mathfrak{R}_q$). Thus the correspondence $\bar{x}^{\alpha^2} = x^{\alpha^2} + t^{\alpha^2}$ mentioned above must be isometric correspondences between the various R_p. This entails that $g_{\alpha^1\beta^1}$ be independent of x^{α^2}. If, further, coördinates x^{α^1} in the R_p and x^a in $\mathfrak{R}_p$ are so chosen that points which correspond to each other in an isometric correspondence have the same coördinates numerically, it follows that $g_{\alpha^1\beta^1} = g_{ab}$. Similarly, we require $g_{\alpha^2\beta^2} = g_{ij}$. These properties are invariant under change of code.

If the components $g_{\alpha^1\alpha^2}$ vanish in one code (and therefore $g^{\alpha^1\alpha^2} = 0$), then $g_{\alpha^1\alpha^2} = 0$ in any code. The special product R_r with this important property is called the direct product of $\mathfrak{R}_p$ and $\mathfrak{R}_q$. §§2–7 are devoted to the direct product, which is by far the most interesting product.[2]

If the components $g_{\alpha^1\alpha^2}$ are unrestricted, except for the requirement that $g_{\alpha\beta}$ be symmetric and positive definite, R_r is called a general product of $\mathfrak{R}_p$ and $\mathfrak{R}_q$.

II. The Direct Product

2. Product-tensors

The components of a tensor fall into classes according to the superscripts (1 or 2) of their indices.[3] The set of components selected from $T^{\alpha\cdots\beta}_{\gamma\cdots\delta}$ by assigning a superscript to each of its indices is called a member of T. If each superscript is 1 (or 2), the member is called the first (or second) member; otherwise, the member is called mixed.

Let the law of transformation for T be

$$\bar{T}^{\alpha_1\cdots\alpha_s}_{\beta_1\cdots\beta_t} = T^{\gamma_1\cdots\gamma_s}_{\delta_1\cdots\delta_t}\frac{\partial\bar{x}^{\alpha_1}}{\partial x^{\gamma_1}}\cdots\frac{\partial\bar{x}^{\alpha_s}}{\partial x^{\gamma_s}}\frac{\partial x^{\delta_1}}{\partial\bar{x}^{\beta_1}}\cdots\frac{\partial x^{\delta_t}}{\partial\bar{x}^{\beta_t}}.$$

If particular superscripts are assigned to the free indices α and β, the quantities on the left side of this equation belong to a member of T. If $x \leftrightarrow \bar{x}$ is a change of code, it is seen that each γ and δ may be assigned the same superscript as the corresponding α and β, showing that this member of T behaves, under a change of

---

[2] The chief results already known are conveniently summarized by W. Mayer in his *Lehrbuch der Differentialgeometrie*, Bd. II, pp. 147–152. Cf. Eisenhart, "Symmetric Tensors of the Second Order whose First Covariant Derivatives are Zero," *Trans. Amer. Math. Soc.*, vol. 25 (1923), pp. 297–306; and Levy, "Symmetric Tensors of the Second Order whose Covariant Derivatives Vanish," *Annals of Math.*, vol. 27 (1926), pp. 91–98.

[3] The operation of assigning a superscript to an index can be given explicit expression by means of the tensors $M^\alpha_\beta = \begin{pmatrix}\delta^{\alpha^1}_{\beta^1} & 0\\ 0 & 0\end{pmatrix}$ and $N^\alpha_\beta = \begin{pmatrix}0 & 0\\ 0 & \delta^{\alpha^2}_{\beta^2}\end{pmatrix}$. For example, $T^{\alpha^1} = M^\alpha_\beta T^\beta$.

code, as a tensor of the same type as T. This is true of each member of T. In this sense, every tensor is compound.

If a member of T vanishes in one code, it therefore vanishes in any code. If every mixed member of T vanishes in a code, T is said to be *breakable*.

If a tensor T is breakable and the first and second members of T depend, in any code, only on variables of the first and second kinds respectively, then T will be called a *product-tensor*. $g_{\alpha\beta}$ and $g^{\alpha\beta}$ are product-tensors; also, $g^{\alpha^1\beta^1} = g^{ab}$ and $g^{\alpha^2\beta^2} = g^{ij}$. The sum, product, and contracted product of product-tensors are product-tensors.

The Cristoffel symbols

$$[\alpha\beta, \gamma] = \frac{1}{2}\left(\frac{\partial g_{\gamma\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\right)$$

and

$$\Gamma^\delta_{\alpha\beta} = g^{\delta\gamma}[\alpha\beta, \gamma]$$

also have the product character. That is: they vanish unless all indices are alike; if all indices are alike, they depend only on variables with indices of the same kind. Also,

$$[\alpha^1\beta^1, \gamma^1] = [ab, c], \qquad [\alpha^2\beta^2, \gamma^2] = [ij, k],$$
$$\Gamma^{\delta^1}_{\alpha^1\beta^1} = \Gamma^d_{ab}, \qquad \Gamma^{\delta^2}_{\alpha^2\beta^2} = \Gamma^l_{ij}.$$

From these properties of $g_{\alpha\beta}$ and $\Gamma^\gamma_{\alpha\beta}$, it follows that the tensors

$$R^\alpha_{\beta\gamma\delta} \equiv \frac{\partial \Gamma^\alpha_{\beta\delta}}{\partial x^\gamma} - \frac{\partial \Gamma^\alpha_{\beta\gamma}}{\partial x^\delta} + \Gamma^\epsilon_{\beta\delta}\Gamma^\alpha_{\epsilon\gamma} - \Gamma^\epsilon_{\beta\gamma}\Gamma^\alpha_{\epsilon\delta}$$

$$R_{\alpha\beta\gamma\delta} \equiv g_{\alpha\epsilon}R^\epsilon_{\beta\gamma\delta}$$

$$R_{\beta\gamma} \equiv R^\alpha_{\beta\gamma\alpha}$$

are product-tensors, and that their first and second members coincide with the respective tensors on $\mathfrak{R}_p$ and $\mathfrak{R}_q$. Also,

$$R \equiv g^{\alpha\beta}R_{\alpha\beta} \equiv g^{\alpha^1\beta^1}R_{\alpha^1\beta^1} + g^{\alpha^2\beta^2}R_{\alpha^2\beta^2} \equiv R_1 + R_2.$$

The covariant derivative of $T^{\alpha_1\cdots\alpha_s}_{\beta_1\cdots\beta_t}$ is

$$T^{\alpha_1\cdots\alpha_s}_{\beta_1\cdots\beta_t,\gamma} \equiv \frac{\partial T^{\alpha_1\cdots\alpha_s}_{\beta_1\cdots\beta_t}}{\partial x^\gamma}$$
$$+ \sum_{\nu=1}^{s} T^{\alpha_1\cdots\alpha_{\nu-1}\delta\alpha_{\nu+1}\cdots\alpha_s}_{\beta_1\cdots\beta_t}\Gamma^{\alpha_\nu}_{\delta\gamma} - \sum_{\nu=1}^{t} T^{\alpha_1\cdots\alpha_s}_{\beta_1\cdots\beta_{\nu-1}\delta\beta_{\nu+1}\cdots\beta_t}\Gamma^\delta_{\beta_\nu\gamma}.$$

For any tensor, it follows that

$$(2.2) \qquad T^{\alpha^1_1\cdots\alpha^1_s}_{\beta^1_1\cdots\beta^1_t,\gamma^2} = \frac{\partial T^{\alpha^1_1\cdots\alpha^1_s}_{\beta^1_1\cdots\beta^1_t}}{\partial x^{\gamma^2}}, \qquad T^{\alpha^2_1\cdots\alpha^2_s}_{\beta^2_1\cdots\beta^2_t,\gamma^1} = \frac{\partial T^{\alpha^2_1\cdots\alpha^2_s}_{\beta^2_1\cdots\beta^2_t}}{\partial x^{\gamma^1}}.$$

If T is a product-tensor the right member of each of these equations vanishes, the remaining mixed members of $T_{,\alpha}$ vanish because T and $\Gamma^{\alpha}_{\beta\gamma}$ are breakable, and the first (second) member of $T_{,\alpha}$ depends only on variables of the first (second) kind. Hence

THEOREM 2.1: *If T is a product-tensor, so is $T_{,\alpha}$.*

Thus the covariant derivatives of the curvature tensor are product-tensors. Their first and second members coincide with the respective tensors on $\mathfrak{R}_p$ and $\mathfrak{R}_q$.

If $T_{,\alpha}$ is breakable, (2.2) shows that the first (second) member of T depends only on variables of the first (second) kind, so that $T_{,\alpha}$ has the same property, and is therefore a product-tensor.

THEOREM 2.2: *If $T_{,\alpha}$ is breakable, it is a product-tensor.*

Since a vector V has no mixed components, the argument just given proves that, if $V_{,\alpha}$ is breakable, V is a product-vector. Combining this with Theorem 2.1, we have

THEOREM 2.3: *V is a product-vector if and only if $V_{,\alpha}$ is a product-tensor.*

The product ζ^{α} of the vectors ξ^{a} of $\mathfrak{R}_p$ and η^{i} of $\mathfrak{R}_q$ is $\zeta^{\alpha} = (\xi^{\alpha^1}, \eta^{\alpha^2})$. Similarly, the product of ξ_a and η_i is $\zeta_\alpha = (\zeta_{\alpha^1}, \eta_{\alpha^2})$. If $\xi_a = g_{ab}\xi^{b}$ and $\eta_i = g_{ij}\eta^{j}$, then, since $g_{\alpha^1\alpha^2} = 0$, their product is connected with the product of ξ^{a} and η^{i} by the relations $\zeta_\alpha = g_{\alpha\beta}\zeta^{\beta}$, $\zeta^{\alpha} = g^{\alpha\beta}\zeta_\beta$.

Any vector $\zeta^{\alpha} = (\zeta^{\alpha^1}, \zeta^{\alpha^2})$ of R_r is the sum of $\zeta_1^{\alpha} = (\zeta^{\alpha^1}, 0)$ in the R_p and $\zeta_2^{\alpha} = (0, \zeta^{\alpha^2})$ in the R_q. Since $g_{\alpha^1\alpha^2} = 0$, ζ_1^{α} and ζ_2^{α} are orthogonal, so that ζ^{α} has squared length $|\zeta|^2 = |\zeta_1|^2 + |\zeta_2|^2$.

If $\theta_{\xi\eta}$ denotes the angle in R_r between a pair of arbitrary unit vectors ξ^{α} and η^{α} of R_r, ϕ_ξ the angle between ξ and ξ_1, ϕ_η the angle between η and η_1, θ_1 the angle between ξ_1 and η_1, and θ_2 the angle between ξ_2 and η_2, then

$$\begin{aligned}\cos\theta_{\xi\eta} &= g_{\alpha\beta}\xi^{\alpha}\eta^{\beta} = g_{\alpha^1\beta^1}\xi^{\alpha^1}\eta^{\beta^1} + g_{\alpha^2\beta^2}\xi^{\alpha^2}\eta^{\beta^2}\\ &= |\xi_1||\eta_1|\cos\theta_1 + |\xi_2||\eta_2|\cos\theta_2\\ &= \cos\phi_\xi\cos\phi_\eta\cos\theta_1 + \sin\phi_\xi\sin\phi_\eta\cos\theta_2.\end{aligned}$$

3. General Theorems

Before examining the detailed structure of product-spaces, we prove five theorems of a general nature.

A Riemann space with metric tensor g_{ab} is *flat* (*Euclidean*) if there exists a coördinate system in which $g_{ab} = \delta_{ab}$. The product of two flat spaces is obviously flat.

THEOREM 3.1: *A space of constant non-vanishing curvature cannot be a product-space.*

If R_r were to have constant curvature $K \neq 0$, then $R_{\alpha\beta\gamma\delta} = KT_{\alpha\beta\gamma\delta}$, where $T_{\alpha\beta\gamma\delta} \equiv g_{\alpha\gamma}g_{\beta\delta} - g_{\beta\gamma}g_{\alpha\delta}$. But $T_{\alpha^1\beta^2\gamma^1\delta^2} = g_{\alpha^1\gamma^1}g_{\beta^2\delta^2}$, which cannot vanish for every choice of its indices, so that the non-breakable tensor T would be proportional to the breakable tensor R. This contradiction proves the theorem.

THEOREM 3.2: *A space $\bar{R}_r$ in geodesic correspondence with a product-space R_r is not necessarily a product-space.*

A counter-example is furnished by a sphere, which is mapped geodesically by central projection on a flat plane tangent to it.

THEOREM 3.3: *$R_r = \mathfrak{R}_p \times \mathfrak{R}_q$ is conformally flat and if and only if:*

a) *when $p = 1$, $q \geqq 1$, $\mathfrak{R}_q$ is of constant curvature;*

b) *when $p \geqq 2$, $q \geqq 2$, $\mathfrak{R}_p$ and $\mathfrak{R}_q$ have constant curvatures for which $K_1 + K_2 = 0$.*

If $p = 1$, $\mathfrak{R}_p$ is flat. If $q = 1$ and $r = 2$, $\mathfrak{R}_q$ and R_r are also flat, and the theorem holds.

If $p = 1$, $q = 2$, $r = 3$, R_r is conformally flat if and only if[4]

$$R_{\alpha\beta\gamma} \equiv R_{\alpha\beta,\gamma} - R_{\alpha\gamma,\beta} + \frac{1}{2(r-1)}(g_{\alpha\gamma}R_{,\beta} - g_{\alpha\beta}R_{,\gamma}) = 0.$$

Since $R_1 = 0$ and $R = R_1 + R_2$, it follows from $2(r-1)R_{\alpha^1\beta^1\gamma^2} = -g_{\alpha^1\beta^1}R_{,\gamma^2} = 0$ that R_2 is also constant, and the condition reduces to $R_{\alpha\beta,\gamma} - R_{\alpha\gamma,\beta} = 0$. In $\mathfrak{R}_q$ we choose orthogonal coördinates, in which $g^{12} = 0$, and have $R_{12} = g^{21}R_{2121} = 0$, whence $R_{\alpha^2\beta^2\gamma^2} = 0$ reduces to $R_{11,2} = 0 = R_{22,1}$. The first of these yields

$$\frac{\partial(g^{22}R_{2112})}{\partial x^2} - 2g^{22}R_{2112}\Gamma^1_{12} = g^{22}\frac{\partial R_{2112}}{\partial x^2} - 2g^{22}R_{2112}(\Gamma^1_{12} + \Gamma^2_{22}) = 0.$$

Cancelling g^{22}, we have

$$\frac{\partial R_{2112}}{\partial x^2} - \frac{R_{2112}}{g}\frac{\partial g}{\partial x^2} = 0,$$

which says that R_{2112}/g is independent of x^2; similarly, this ratio is independent of x^1. Thus $R_{2112} = -Kg_{22}g_{11}$, which is the condition in orthogonal coördinates that $\mathfrak{R}_q$ have constant curvature.

If $r \geqq 4$, R_r will be conformally flat if and only if the conformal curvature tensor vanishes, that is,

$$C_{\alpha\beta\gamma\delta} \equiv R_{\alpha\beta\gamma\delta} + \frac{1}{r-2}T_{\alpha\beta\gamma\delta} + \frac{R}{(r-1)(r-2)}U_{\alpha\beta\gamma\delta} = 0, \tag{3.1}$$

where $\begin{cases} T_{\alpha\beta\gamma\delta} \equiv g_{\alpha\gamma}R_{\beta\delta} - g_{\alpha\delta}R_{\beta\gamma} + g_{\beta\delta}R_{\alpha\gamma} - g_{\beta\gamma}R_{\alpha\delta}, \\ U_{\alpha\beta\gamma\delta} \equiv g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta}. \end{cases}$

Necessary conditions are found first. $C_{\alpha^1\beta^2\gamma^1\delta^2} = 0$ gives

$$(r-1)(g_{\alpha^1\gamma^1}R_{\beta^2\delta^2} + g_{\beta^2\delta^2}R_{\alpha^1\gamma^1}) + (R_1 + R_2)(-g_{\alpha^1\gamma^1}g_{\beta^2\delta^2}) = 0. \tag{3.2}$$

From this it follows, on multiplication by $g^{\beta^2\delta^2}$, that

$$R_{\alpha^1\gamma^1} = \frac{qR_1 - (p-1)R_2}{q(r-1)}g_{\alpha^1\gamma^1}, \tag{3.3}$$

[4] Cf. R.G., p. 92.

and, on multiplication by $g^{\alpha^1\gamma^1}$, that

$$R_{\beta^2\delta^2} = \frac{pR_2 - (q-1)R_1}{p(r-1)} g_{\beta^2\delta^2}. \tag{3.4}$$

If $p = 1$, $q \geqq 3$, then $R_1 = 0$, $r - 1 = q$, and (3.4) shows that

$$R_{\beta^2\delta^2} = \frac{R_2}{q} g_{\beta^2\delta^2}.$$

Thus $\mathfrak{R}_q$ is an Einstein space; since $q > 2$, R_2 is a constant. Substitution in $C_{\alpha^2\beta^2\gamma^2\delta^2} = 0$ gives

$$R_{\alpha^2\beta^2\gamma^2\delta^2} = \frac{R_2}{q(1-q)} (g_{\alpha^2\gamma^2} g_{\beta^2\delta^2} - g_{\alpha^2\delta^2} g_{\beta^2\gamma^2}),$$

and $\mathfrak{R}_q$ has constant curvature. By substitution in (3.1) this necessary condition is seen to be also sufficient. This completes the proof of a).

If $p \geqq 2$, and $q \geqq 2$, either (3.3) or (3.4) shows that

$$p(p-1)R_2 + q(q-1)R_1 = 0, \tag{3.5}$$

and differentiation of (3.5) shows that R_1 and R_2 are constants. (3.5) applied to (3.3) and (3.4) show that

$$pR_{\alpha^1\gamma^1} = R_1 g_{\alpha^1\gamma^1}, \qquad qR_{\beta^2\delta^2} = R_2 g_{\beta^2\delta^2}; \tag{3.6}$$

that is, each factor space is an Einstein space. (3.5) and (3.6) are necessary and sufficient conditions that $C_{\alpha\beta\gamma\delta}$ be breakable.

When (3.6) is used to simplify the conditions $C_{\alpha^1\beta^1\gamma^1\delta^1} = 0$, the result is

$$R_{\alpha^1\beta^1\gamma^1\delta^1} - \frac{1}{r-2}\left(\frac{2R_1}{p} - \frac{R_1 + R_2}{r-1}\right) U_{\alpha^1\beta^1\gamma^1\delta^1} = 0. \tag{3.7}$$

Substitution from (3.5) into (3.7) gives

$$R_{\alpha^1\beta^1\gamma^1\delta^1} = \frac{R_1}{p(1-p)} U_{\alpha^1\beta^1\gamma^1\delta^1}, \tag{3.8}$$

showing that $\mathfrak{R}_p$ is a space of constant curvature $K_1 = R_1/p(1-p)$. Similarly, $\mathfrak{R}_q$ is a space of constant curvature $K_2 = R_2/q(1-q)$. (3.5) is the same as $K_1 + K_2 = 0$. The conditions stated in [3.3] are thus shown to be necessary. Direct substitution in (3.1) shows them to be sufficient. q.e.d.

It seems remarkable that this theorem lays no further restriction on the dimensions of the factor-spaces.

Theorem 3.4: *A product-space is an Einstein space if and only if each factor-space is an Einstein space and*

$$\frac{R_1}{p} = \frac{R_2}{q}.$$

For R_r to be an Einstein space it is necessary and sufficient that

$$R_{\alpha\beta} = \frac{R}{r} g_{\alpha\beta} = \frac{R_1 + R_2}{p+q} g_{\alpha\beta}.$$

Since $R_{\alpha\beta}$ is a product-tensor, and $R_1/p = R_2/q$ implies that either equals $(R_1 + R_2)/(p + q)$, the sufficiency of the condition is evident.

It is necessary that

$$R_{\alpha^1\beta^1} = \frac{R_1 + R_2}{p+q} g_{\alpha^1\beta^1}.$$

Contraction with $g^{\alpha^1\beta^1}$ shows that $R_1/p = R_2/q$, whence

$$R_{\alpha^1\beta^1} = \frac{R_1}{p} g_{\alpha^1\beta^1}$$

and $\mathfrak{R}_p$ is an Einstein space. Similarly for $\mathfrak{R}_q$. q.e.d.

If $\mathfrak{R}_p$ can be imbedded in a Euclidean $\mathfrak{E}_{p+p'}$ but not in an $\mathfrak{E}_{p+n}$ if $n < p'$, then $\mathfrak{R}_p$ is said to be of *class* p'.

Let $\mathfrak{R}_p$, $\mathfrak{R}_q$, and R_r have classes p', q', and r'. Let $\bar{p} = p + p'$, $\bar{q} = q + q'$, define $\bar{r} = \bar{p} + \bar{q}$, and, for the remainder of this section, let indices have the following ranges: $i^1 = 1, \cdots, \bar{p}$; $i^2 = \bar{p} + 1, \cdots, \bar{r}$; $i = 1, \cdots, \bar{r}$; $\alpha^1 = 1, \cdots, p$; $\alpha^2 = \bar{p} + 1, \cdots, \bar{p} + q$; $\alpha = 1, \cdots, p, \bar{p} + 1, \cdots, \bar{p} + q$.

If

$$\begin{cases} y^{i^1} = f^{i^1}(x^{\alpha^1}), \\ y^{i^2} = f^{i^2}(x^{\alpha^2}), \end{cases} \tag{3.8}$$

define imbeddings of $\mathfrak{R}_p$ in an $\mathfrak{E}_{\bar{p}}$ and of $\mathfrak{R}_q$ in an $\mathfrak{E}_{\bar{q}}$, then $y^i = f^i(x^\alpha)$ define an imbedding of R_r in an $\mathfrak{E}_{\bar{r}}$.[5] Hence $r' \leqq p' + q'$. The possibility $r' < p' + q'$ is excluded by

THEOREM 3.5: $r' = p' + q'$; *that is, the class of a direct product-space is the sum of the classes of the factor-spaces.*

If $r' < p' + q'$, equations

$$z^{i'} = \phi^{i'}(x^\alpha) \qquad (i' = 1, \cdots, r + r') \tag{3.9}$$

exist which imbed R_r in an $\mathfrak{E}_{r+r'}$. We regard $\mathfrak{E}_{r+r'}$ as a subspace

$$z^\nu = c^\nu \qquad (\nu = r + r' + 1, \cdots, \bar{r}) \tag{3.10}$$

of $\mathfrak{E}_{\bar{r}}$. There then exists an isometric point transformation of $\mathfrak{E}_{\bar{r}}$ carrying the subspace defined by (3.8) into the subspace defined by (3.9) and (3.10). Every isometric point transformation of $\mathfrak{E}_{\bar{r}}$ preserves the quadratic form $ds^2 = \sum_1^{\bar{r}} (dy^i)^2$, and is therefore of the form $\bar{y}^i = a^i_j y^j + a^i$ with a^i_j orthogonal. Let

[5] It follows from the equations $g_{\alpha\beta} = \dfrac{\partial y^i}{\partial x^\alpha}\dfrac{\partial y^i}{\partial x^\beta}$ for the imbedding of any R_r in an $\mathfrak{E}_{\bar{r}}$ that an R_r which has an imbedding of the form (3.8) is a direct product-space.

$z^i = a^i_j y^j + a^i$ be such a transformation which carries (3.8) into (3.9–10). Then $a^\nu_j y^j + a^\nu = c^\nu$. Differentiation of this equation shows by (3.8) that

$$(3.11) \qquad a^\nu_{j^1} \frac{\partial y^{j^1}}{\partial x^{\alpha^1}} = 0, \qquad a^\nu_{j^2} \frac{\partial y^{j^2}}{\partial x^{\alpha^2}} = 0.$$

Since $\| a^\nu_j \|$ has rank $\bar{r} - (r + r') > 0$, either $\| a^\nu_{j^1} \|$ or $\| a^\nu_{j^2} \|$ has positive rank, and it follows from (3.11) that there is then a functional relation either between the y^{i^1} or between the y^{i^2}, i.e., that either $\mathfrak{R}_p$ is of class $< p'$, or $\mathfrak{R}_q$ is of class $< q'$. This contradiction proves the theorem.

4. Geodesic Subspaces

In this section we define products and projections of subspaces and study their geodesic properties.

Let $x^a = x^a(u^{a^3})$ $(a^3 = 1, \cdots, l \leqq p)$ and $x^i = x^i(u^{i^3})$ $(i^3 = 1, \cdots, m \leqq q)$ define subspaces $\mathfrak{R}_l$ and $\mathfrak{R}_m$ of $\mathfrak{R}_p$ and $\mathfrak{R}_q$. The equations $x^{\alpha^1} = x^{\alpha^1}(u^{a^3})$, $x^{\alpha^2} = x^{\alpha^2}(u^{i^3})$ define a subspace R_n $(n = l + m)$ of R_r which is called the product of $\mathfrak{R}_l$ and $\mathfrak{R}_m$.

Let $x^\alpha = x^\alpha(u^{\alpha^3})$ $(\alpha^3 = 1, \cdots, n < r)$ be a subspace R_n of R_r, and let $x^{\alpha^2} = c^{\alpha^2}$ define an R_p. If $\| \partial x^{\alpha^2}/\partial u^{\alpha^3} \|$ has rank ρ when $x^{\alpha^2} = c^{\alpha^2}$, the equations $c^{\alpha^2} = x^{\alpha^2}(u^{\alpha^3})$ can be solved for ρ of the u^{α^3} in terms of the remaining $n - \rho$ parameters, which we call w^{α^4} $(\alpha^4 = 1, \cdots, n - \rho)$. By substituting these ρ u's in the equations $x^{\alpha^1} = x^{\alpha^1}(u^{\alpha^3}) = x^{\alpha^1}(w^{\alpha^4})$, we find the intersection of R_n with R_p; if, after the substitution, $\| \partial x^{\alpha^1}/\partial w^{\alpha^4} \|$ has rank l, the intersection is of dimension l. The intersection of R_n with R_q is defined similarly.

The projection of R_n on R_p is more useful for our purposes. It is the subspace of R_p defined by the equations

$$\begin{cases} x^{\alpha^1} = x^{\alpha^1}(u^{\alpha^3}) \\ x^{\alpha^2} = c^{\alpha^2} \end{cases}$$

The dimension of the projection equals the rank of $\| \partial x^{\alpha^1}/\partial u^{\alpha^3} \|$ and does not exceed the lesser of p and n. The projection of R_n on R_q is defined similarly. The projection of R_n on R_p contains the intersection of R_n with R_p. If R_n is a product-subspace, its intersections with the factor-spaces coincide with its projections on the factor-spaces, and R_n is their product.

If $P(t)$ is an arc defined by $x^\alpha = x^\alpha(t)$ $(0 \leqq t \leqq 1)$, its projections are $P_1(t)$: $x^a = x^a(t)$ and $P_2(t)$: $x^i = x^i(t)$. The length of $P(t)$ is defined formally by

$$s = \int_0^1 \frac{ds}{dt}\, dt = \int_0^1 \left[\left(\frac{ds_1}{dt}\right)^2 + \left(\frac{ds_2}{dt}\right)^2 \right]^{\frac{1}{2}} dt.$$

If s exists, then evidently the lengths $s_\nu = \int_0^1 \frac{ds_\nu}{dt}\, dt$ $(\nu = 1, 2)$ of the projections $P_\nu(t)$ both exist, for $s_\nu \leqq s$. If both s_ν exist, then

$$s_1 + s_2 = \int_0^1 \left(\frac{ds_1}{dt} + \frac{ds_2}{dt} \right) dt = \int_0^1 \left[\left(\frac{ds_1}{dt}\right)^2 + \left(\frac{ds_2}{dt}\right)^2 + 2 \frac{ds_1}{dt} \frac{ds_2}{dt} \right]^{\frac{1}{2}} dt;$$

since g_{ab} and g_{ij} are positive definite, $\dfrac{ds_1}{dt}\dfrac{ds_2}{dt} \geqq 0$, and $s_1 + s_2 \geqq s$, so that s also exists. This shows that

THEOREM 4.1: *An arc in R_r is rectifiable if and only if each of its projections is rectifiable.*

Now let $P(t)$ be a geodesic of R_r, and for any u let $\dot{u} \equiv du/dt$. The $x^\alpha(t)$ satisfy the equations

$$T^{\alpha\beta} \equiv \dot{x}^\alpha(\ddot{x}^\beta + \Gamma^\beta_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta) - \dot{x}^\beta(\ddot{x}^\alpha + \Gamma^\alpha_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta) = 0.$$

Of these, the equations $T^{\alpha^1\beta^1} = 0$ and $T^{\alpha^2\beta^2} = 0$ show that the projections $P_1(t)$ in $\mathfrak{R}_p$ and $P_2(t)$ in $\mathfrak{R}_q$ are geodesics.

Suppose conversely that $P_1(s_1)$ and $P_2(s_2)$ are geodesics of arc-lengths s_1 and s_2 given by $x^a = x^a(s_1)$ and $x^i = x^i(s_2)$ in $\mathfrak{R}_p$ and $\mathfrak{R}_q$. We seek a function $s_2 = f(s_1)$ such that $x^\alpha(s_1) = (x^a(s_1), x^i(f(s_1)))$ shall move along a geodesic $P(s_1)$ lying in R_r on the product $P(s_1, s_2) = (P_1(s_1), P_2(s_2))$ of $P_1(s_1)$ and $P_2(s_2)$. Since s_1 may not be the arc-length of $x^\alpha(s_1)$, we use the condition $T^{\alpha\beta} = 0$. $T^{\alpha^1\beta^1} = 0$ is satisfied by hypothesis. $T^{\alpha^1\beta^2} = 0$ gives

$$\frac{dx^{\alpha^1}}{ds_1}\left(\frac{d^2x^{\beta^2}}{ds_2^2} + \Gamma^{\beta^2}_{\gamma^2\delta^2}\frac{dx^{\gamma^2}}{ds_2}\frac{dx^{\delta^2}}{ds_2}\right)\left(\frac{ds_2}{ds_1}\right)^2 + \frac{dx^{\alpha^1}}{ds_1}\frac{dx^{\beta^2}}{ds_2}\frac{d^2s_2}{ds_1^2}$$
$$- \frac{dx^{\beta^2}}{ds_2}\left(\frac{d^2x^{\alpha^1}}{ds_1^2} + \Gamma^{\alpha^1}_{\beta^1\gamma^1}\frac{dx^{\beta^1}}{ds_1}\frac{dx^{\gamma^1}}{ds_1}\right) = 0.$$

Since s_1 and s_2 are arc-lengths by hypothesis, the terms in parentheses vanish. Since $\dfrac{dx^{\alpha^1}}{ds_1}\dfrac{dx^{\beta^2}}{ds_2}$ cannot vanish for all choices of α^1 and β^2, it follows that $\dfrac{d^2s_2}{ds_1^2} = 0$, and $s_2 = as_1 + b$. Direct inspection shows that this condition is sufficient for $T^{\alpha^2\beta^2} = 0$ also to be satisfied by $P(s_1)$. This completes the proof of

THEOREM 4.2: *A curve is a geodesic if and only if each of its projections is a geodesic. A curve on the product of two geodesics, with arc-lengths s_1 and s_2, is a geodesic if and only if it has an equation of the form $s_2 = as_1 + b$.*

We shall say that a subspace of a Riemann space $\mathfrak{R}_p$ is plane (or geodesic) at a point P if it contains entirely each geodesic of $\mathfrak{R}_p$ which is tangent to it at P. A subspace is a plane (or is totally geodesic) if it is plane (or geodesic) at each of its points.[6]

It is a corollary of [4.2] that each $\mathfrak{R}_p$ and each $\mathfrak{R}_q$ is a plane (is totally geodesic) in R_r. We examine other subspaces of R_r for this property.

Normal coördinates are useful. Let $P = (P_1, P_2)$, let y^a and y^i be normal coördinates at P_1 in $\mathfrak{R}_p$ and at P_2 in $\mathfrak{R}_q$, and consider the code which is the product of these coördinate systems. A geodesic C of R_r issuing from P has projections which are geodesics ([4.2]), and which therefore have equations

[6] These definitions are from Cartan, *Leçons sur la géométrie des espaces de Riemann*, p. 131.

$y^a = \xi^a t$ and $y^i = \xi^i t$. C then has equations $y^\alpha = \xi^\alpha t$ and ξ^α is tangent to C at P. This proves

THEOREM 4.3: *The product of two normal coördinate systems is a normal code.*

Now let $\mathfrak{R}_p$ be any Riemann space with normal coördinates y^a at a point P_1, and let ξ^a_b be any p independent vectors at P_1. The equations $y^a = z^{b^3}\xi^a_{b^3}$ $(b^3 = 1, \cdots, l \leqq p)$ define a subspace $\mathfrak{R}_l$ which contains entirely any geodesic tangent to it at P_1—in other words, $\mathfrak{R}_l$ is plane at P_1. If, conversely, $\mathfrak{R}_l$ is plane at P_1, we use normal coördinates at P_1, choose l vectors tangent to $\mathfrak{R}_l$ at P_1, and write its equations in the form $y^a = z^{b^3}\xi^a_{b^3}$. Thus a subspace $\mathfrak{R}_l$ is plane at P_1 if and only if, when y^a are normal coördinates at P_1, and $\xi^a_{a^3}$ $(a^3 = 1, \cdots, l)$ are vectors tangent to $\mathfrak{R}_l$ at P_1, it has parametric equations $y^a = z^{a^3}\xi^a_{a^3}$.

At a point $P = (P_1, P_2)$ in R_r we now use the normal code y^α which is the product of normal coördinate systems y^a at P_1 and y^i at P_2. If a subspace R_n is plane at P, it has parametric equations $y^\alpha = z^{\alpha^3}\xi^\alpha_{\alpha^3}$ $(\alpha^3 = 1, \cdots, n)$, where $\xi^\alpha_{\alpha^3}$ are the vectors tangent to it at P. Its first projection $\mathfrak{R}_l$ is $y^a = z^{\alpha^3}\xi^a_{\alpha^3}$. The Jacobian $\| \partial y^a/\partial z^{\alpha^3} \| = \| \xi^a_{\alpha^3} \|$ has rank l; suppose that $\| \xi^a_{a^3} \|$ $(a^3 = 1, \cdots, l)$ has rank l, and let $a^4 = l + 1, \cdots, n$. Then $\xi^a_{a^4} = c^{a^3}_{a^4}\xi^a_{a^3}$ and $y^a = (z^{a^3} + z^{a^4}c^{a^3}_{a^4})\xi^a_{a^3}$. Since the $\xi^a_{a^3}$ are tangent to $\mathfrak{R}_l$, this shows that $\mathfrak{R}_l$ is plane at P_1. Similarly, the other projection $\mathfrak{R}_m$ is plane at P_2.

Conversely, let $\mathfrak{R}_l$ and $\mathfrak{R}_m$ be plane at P_1 and P_2. By [4.2], a geodesic C issuing from P tangent to R_n has projections which are geodesics issuing from P_1 and P_2 tangent to $\mathfrak{R}_l$ and $\mathfrak{R}_m$. Since these projections lie entirely in $\mathfrak{R}_l$ and $\mathfrak{R}_m$, C lies entirely in R_n, so that R_n is plane at P. This completes the proof of

THEOREM 4.4: *A subspace is plane at $P = (P_1, P_2)$ if and only if its projections are plane at P_1 and P_2.*

Since a product-subspace is the product of its projections, we see that

COROLLARY 4.41: *A product-subspace is plane at $P = (P_1, P_2)$ if and only if its factors are plane at P_1 and P_2.*

Combined with the definition of a plane, [4.4] gives immediately

THEOREM 4.5: *A subspace is a plane if and only if each of its projections is a plane.*

COROLLARY 4.51: *A product-subspace is a plane if and only if its factors are planes.*

5. Parallel Displacement

ξ^α at x^α and $\xi^\alpha + d\xi^\alpha$ at $x^\alpha + dx^\alpha$ are parallel if and only if $d\xi^\alpha = -\xi^\beta\Gamma^\alpha_{\beta\gamma}\,dx^\gamma$. On any one of these equations, all the indices are of the same kind. To construct a vector $\xi^\alpha + d\xi^\alpha$ at $x^\alpha + dx^\alpha$ parallel in R_r to ξ^α at x^α, we therefore construct vectors $\xi^{\alpha^\nu} + d\xi^{\alpha^\nu}$ $(\nu = 1, 2)$ at $x^{\alpha^\nu} + dx^{\alpha^\nu}$ parallel in the ν^{th} factor-space to ξ^{α^ν} at x^{α^ν}. Then $\xi^\alpha + d\xi^\alpha = (\xi^{\alpha^1} + d\xi^{\alpha^1}, \xi^{\alpha^2} + d\xi^{\alpha^2})$ at $x^\alpha + dx^\alpha$ in R_r is the required vector.

Parallelism along an arc has equally simple properties. Let $R_0 = (P_0, Q_0)$,

$R_1 = (P_1, Q_1)$, and let $C(R_0, R_1)$ be an arc given by $x^\alpha = x^\alpha(t)$ $(0 \leqq t \leqq 1)$. All the indices on any one of the equations $\dot{\xi}^\alpha = -\xi^\beta\Gamma^\alpha_{\beta\gamma}\dot{x}^\gamma$ must be of the same kind, and the equations fall into two entirely distinct sets. The displacement of ξ^α by parallelism from R_0 to R_1, may thus be effected by displacing the projection ξ^{α^1} by parallelism along $x^{\alpha^1} = x^{\alpha^1}(t)$ from (P_0, Q_0) to (P_1, Q_0) in the R_p whose points are (P, Q_0), ξ^{α^2} being constant, and then displacing the projection ξ^{α^2} by parallelism along $x^{\alpha^2} = x^{\alpha^2}(t)$ from (P_1, Q_0) to (P_1, Q_1) in the R_q whose points are (P_1, Q), ξ^{α^1} being constant. Conversely, any displacement according to this law is a parallel displacement in R_r. Thus we have

THEOREM 5.1: *A vector is parallel for an infinitesimal displacement, or along an arc, if and only if its projections are parallel for the projections of the infinitesimal displacement or along the projections of the arc.*

From this fact it is easy to show that

THEOREM 5.2: *The holonomic group of R_r is the direct product of the holonomic groups of $\mathfrak{R}_p$ and $\mathfrak{R}_q$.*

The holonomic group[7] at a point P (of any Riemann space) is defined as follows: Let C be any oriented closed curve beginning and ending at P. If a vector ξ^α is carried by parallelism from P once around C and returns to a position $\bar{\xi}^\alpha$, then $\bar{\xi}^\alpha = a^\alpha_\beta(C)\xi^\beta$. These transformations form a group $H(P)$ which is called the holonomic group at P. $H(P)$ and $H(Q)$ are conjugate under the group of non-singular linear transformations, which justifies reference to $H(P)$ for any P as "the" holonomic group of the space.

Now let C be an oriented closed curve beginning and ending at $P = (P_1, P_2)$ in our product space R_r, and let C_1 and C_2 be its projections. [5.1] shows that $\bar{\xi}^\alpha = a^\alpha_\beta(C)\xi^\beta$, where $a^{\alpha^\nu}_{\beta^\nu}(C) = a^{\alpha^\nu}_{\beta^\nu}(C_\nu)$ $(\nu = 1, 2)$, and $a^{\alpha^1}_{\beta^2} = 0 = a^{\alpha^2}_{\beta^1}$. This proves [5.2].

We return to infinitesimal parallelism and study the curvature of R_r. If dx^α and δx^α are differentials at x^α, and ξ^α is carried by parallelism from x^α around the parallelogram $(dx^\alpha, \delta x^\alpha)$, it suffers an increment $\nabla\xi^\alpha = \xi^\beta R^\alpha_{\beta\gamma\delta}\, dx^\gamma\, \delta x^\delta = (\nabla\xi^{\alpha^1}, \nabla\xi^{\alpha^2})$. $\nabla\xi^{\alpha^\nu} = \xi^{\beta^\nu} R^{\alpha^\nu}_{\beta^\nu\gamma^\nu\delta^\nu}\, dx^{\gamma^\nu}\, \delta x^{\delta^\nu}$ $(\nu = 1, 2)$ is the result of displacing ξ^{α^ν} around a parallelogram $(dx^{\alpha^\nu}, \delta x^{\alpha^\nu})$ in the ν^{th} factor-space.

Letting $\xi^\alpha = \delta x^\alpha$, resolving $\nabla\xi^\alpha$ along dx^α, and normalizing the resulting scalar, we have the curvature of R_r for the two-spread $(dx^\alpha, \delta x^\alpha)$:

$$K = \frac{R_{\alpha\beta\gamma\delta}\, dx^\alpha\, \delta x^\beta\, dx^\gamma\, \delta x^\delta}{(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma})dx^\alpha\, \delta x^\beta\, dx^\gamma\, \delta x^\delta}. \tag{5.1}$$

In studying this formula the bivector[8] $D^{\alpha\beta} = dx^\alpha\, \delta x^\beta - dx^\beta\, \delta x^\alpha$ is useful. Its squared modulus is $|D|^2 = D^{\alpha\beta}D_{\alpha\beta} = (g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma})\, dx^\alpha\, \delta x^\beta\, dx^\gamma\, \delta x^\delta$. If this expression in (5.1) differs from one, we choose a pair of orthogonal unit vectors dx^α and δx^α in the two-spread (d, δ) and, with the new vectors, we have

[7] Cf. Veblen and Whitehead, *The Foundations of Differential Geometry* (Cambridge Tract No. 29), p. 92.

[8] On bivectors, see Cartan, loc. cit., Chap. I, Sec. II.

$|D|^2 = D^{\alpha\beta}D_{\alpha\beta} = 1$. The projections of this bivector on the factor-spaces are $D^{\alpha^1\beta^1}$ and $D^{\alpha^2\beta^2}$. The angle ϕ between the bivectors $U^{\alpha\beta}$ and $V^{\alpha\beta}$ is given by $|U||V|\cos\phi = U^{\alpha\beta}V_{\alpha\beta}$. Thus $|D_1|^2 = D^{\alpha^1\beta^1}D_{\alpha^1\beta^1} = \cos^2\phi_1$, where ϕ_1 is the angle between $D^{\alpha\beta}$ and $D^{\alpha^1\beta^1}$; similarly for $|D_2|^2$.

With these formulas, (5.1) may be written as

$$K = R_{\alpha\beta\gamma\delta}\,dx^\alpha\,\delta x^\beta\,dx^\gamma\,\delta x^\delta,$$

$$K = R_{\alpha^1\beta^1\gamma^1\delta^1}\,dx^{\alpha^1}\,\delta x^{\beta^1}\,dx^{\gamma^1}\,\delta x^{\delta^1} + R_{\alpha^2\beta^2\gamma^2\delta^2}\,dx^{\alpha^2}\,\delta x^{\beta^2}\,dx^{\gamma^2}\,\delta x^{\delta^2} \tag{5.2}$$

$$K = K_1 D_{\alpha^1\beta^1}D^{\alpha^1\beta^1} + K_2 D_{\alpha^2\beta^2}D^{\alpha^2\beta^2},$$

$$K = K_1\cos^2\phi_1 + K_2\cos^2\phi_2. \tag{5.3}$$

Here K_1 and K_2 are the curvatures of R_p and R_q for the two-spreads associated with $D^{\alpha^1\beta^1}$ and $D^{\alpha^2\beta^2}$. (5.3) is analogous to Euler's formula for normal curvature in classical differential geometry.

There are special cases of (5.3) corresponding to the possible intersections of the two-spread $(dx^\alpha, \delta x^\alpha)$ with the factor-spaces. These intersections are determined by the ranks s_1 and s_2 of the matrices $S_1 = \left\| \begin{matrix} dx^{\alpha^1} \\ \delta x^{\alpha^1} \end{matrix} \right\|$ and $S_2 = \left\| \begin{matrix} dx^{\alpha^2} \\ \delta x^{\alpha^2} \end{matrix} \right\|$. $s_\nu < 2$ means $D^{\alpha^\nu\beta^\nu} = 0$ $(\nu = 1, 2)$. Since $S = \left\| \begin{matrix} dx^{\alpha} \\ \delta x^{\alpha} \end{matrix} \right\|$ has rank 2 by hypothesis, $s_1 + s_2 \geqq 2$. $(dx^\alpha, \delta x^\alpha)$ will have the point P_1 (or P_2), a one-spread, or itself in common with R_p (or R_q) according as s_2 (or s_1) $= 2, 1,$ or 0. Now suppose, for example that $s_1 = 2$ and $s_2 = 1$. Then S_2 may be supposed to have the form $\left\| \begin{matrix} \delta^{\alpha^2}_{p+1} \\ 0 \end{matrix} \right\|$, and it follows from (5.2) that $K = K_1\cos^2\phi_1$ (since $D^{\alpha^2\beta^2} = 0$, K_2 is undefined and (5.3) has no meaning). By considering similarly other values of s_1 and s_2, we verify the facts contained in

THEOREM 5.3: *The ranks s_1 and s_2 of the matrices $S_1 = \left\| \begin{matrix} dx^{\alpha^1} \\ \delta x^{\alpha^1} \end{matrix} \right\|$ and $S_2 = \left\| \begin{matrix} dx^{\alpha^2} \\ \delta x^{\alpha^2} \end{matrix} \right\|$, the intersections of the two-spread $(dx^\alpha, \delta x^\alpha)$ with the factor-spaces, the curvature K of R_r for $(dx^\alpha, \delta x^\alpha)$, the curvatures K_1 and K_2 of $\mathfrak{R}_p$ and $\mathfrak{R}_q$ for $(dx^{\alpha^1}, \delta x^{\alpha^1})$ and $(dx^{\alpha^2}, \delta x^{\alpha^2})$, and the angles ϕ_1 and ϕ_2 in R_r between these two-spreads and $(dx^\alpha, \delta x^\alpha)$ are related according to the following table:*

s_1	s_2	Intersection of $(dx, \delta x)$ with R_p	Intersection of $(dx, \delta x)$ with R_q	K
0	2	P_1	Lies in R_q	K_2
2	0	Lies in R_p	P_2	K_1
1	1	A one-spread	A one-spread	0
1	2	P_1	A one-spread	$K_2\cos^2\phi_2$
2	1	A one-spread	P_2	$K_1\cos^2\phi_1$
2	2	P_1	P_2	$K_1\cos^2\phi_1 + K_2\cos^2\phi_2$

The case $s_1 = 1 = s_2$ gives

COROLLARY 5.31: *A geodesic two-spread which cuts both factor-spaces is flat.*

PROOF: K is defined as the Gaussian curvature of the geodesic two-spread tangent to the two-spread $(dx^\alpha, \delta x^\alpha)$. [4.51] shows that such a two-spread is totally geodesic, and [5.3] shows that $K = 0$ at any point at which it is geodesic, i.e., at each of its points. $K \equiv 0$ means that two-spread is flat.

6. Parallel Fields of Vector-spaces

The set of vectors at a point P linearly dependent on p_1 linearly independent vectors at P is called a vector-space V_{p_1}. If a V_{p_1} is given at each point, we speak of a field of vector-spaces; the symbol V_{p_1} may usually be used without confusion to denote a field of vector-spaces or the space belonging to the field at a given point.

It may happen that any vector of V_{p_1} at any point P, displaced by parallelism from P along any arc, remains a vector of V_{p_1}; in this case V_{p_1} will be called a parallel field of vector-spaces. In our product-space R_r, the field of vector-spaces tangent to the R_p is evidently a parallel field; similarly for the R_q.

Let V_{r_1} be a field of vector-spaces of R_r spanned by ξ^α_σ $(\sigma = 1, \cdots, r_1)$. The projection of V_{r_1} on the subspace R_p given by $x^{\alpha^2} = c^{\alpha^2}$ is a field of vector-spaces $V_{p_1}(x^{\alpha^1})$, where p_1 is the rank of the matrix $\| \xi^{\alpha^1}_\sigma(x^{\alpha^1}, c^{\alpha^2}) \|$; similarly for $V_{q_1}(x^{\alpha^2})$ on the R_q. Suppose now that V_{r_1} is a parallel field of vector-spaces. In the R_p given by $x^{\alpha^2} = c^{\alpha^2}$, let a vector ξ^{α^1} of $V_{p_1}(x^{\alpha^1})$ be displaced by parallelism in R_r from P to P' along an arc lying entirely in R_p. By [5.1], ξ^{α^1} moves also by parallelism in R_p, and ξ^α moves by parallelism in R_r. By hypothesis, ξ^α stays in V_{r_1}, so that ξ^{α^1} stays in V_{p_1}, which is thus shown to be a parallel field of vector-spaces in R_p; similarly for V_{q_1}. If, conversely, each V_{p_1} and each V_{q_1} is a parallel field of vector-spaces, then V_{r_1} is a parallel field of vector-spaces. In other words,

THEOREM 6.1: *A field of vector-spaces in R_r is parallel if and only if each of its projections is parallel.*

If V_{r_1} is the product (direct sum, or join) of V_{p_1} and V_{q_1}, it follows that

COROLLARY 6.11: *The product of two fields of vector-spaces is a parallel field if and only if each factor is parallel.*

An absolutely parallel vector-field is equivalent to a parallel V_1. If ξ^α is absolutely parallel in R_r, then

$$\xi^{\alpha^1}_{,\beta^2} = \frac{\partial \xi^{\alpha^1}}{\partial x^{\beta^2}} = 0, \qquad \xi^{\alpha^2}_{,\beta^1} = \frac{\partial \xi^{\alpha^2}}{\partial x^{\beta^1}} = 0,$$

and ξ^α is a product-vector. This proves

COROLLARY 6.12: *A vector of R_r is parallel if and only if it is the product of two parallel vectors.*

COROLLARY 6.13: *If $\mathfrak{R}_p$ and $\mathfrak{R}_q$ have p_1 and q_1 independent parallel vectors, then R_r has $r_1 = p_1 + q_1$ independent parallel vectors.*

The chief known result on product-spaces is

THEOREM 6.2: *If a Riemann space $\mathfrak{R}_p$ has a parallel field of vector-spaces V_{p_1}, and $p_1 < p$, then $\mathfrak{R}_p$ is the direct product of an $\mathfrak{R}_{p_1}$ and an $\mathfrak{R}_{p-p_1}$.*

This theorem and its proof are to be found in Mayer, loc. cit. The spaces V_{p_1} are tangent to the subspaces R_{p_1}, and the spaces V_{p-p_1} of vectors normal to V_{p_1} are tangent to the subspaces R_{p-p_1} and also form a parallel field.

Let $\mathfrak{R}_p$ have p_1 independent parallel vectors. They span a parallel field V_{p_1} of vector-spaces. We apply [6.2] to $\mathfrak{R}_p$ and then again to $\mathfrak{R}_{p_1}$, in which each parallel field is equivalent to a parallel vector-space V_1, and have

COROLLARY 6.21: *An $\mathfrak{R}_p$ with exactly p_1 independent parallel vectors is the product of a flat $\mathfrak{E}_{p_1}$ and an $\mathfrak{R}_{p-p_1}$ with no parallel vectors.*

This corollary furnishes a useful normal form for a product-space with parallel vectors. Let $\mathfrak{R}_p$ and $\mathfrak{R}_q$ have p_1 and q_1 parallel vectors. Then $\mathfrak{R}_p = \mathfrak{E}_{p_1} \times \mathfrak{R}_{p-p_1}$, $\mathfrak{R}_q = \mathfrak{E}_q \times \mathfrak{R}_{q-q_1}$, and $R_r = (\mathfrak{E}_{p_1} \times \mathfrak{E}_{q_1}) \times (\mathfrak{R}_{p-p_1} \times \mathfrak{R}_{q-q_1}) = \mathfrak{E}_{r_1} \times \mathfrak{R}_{r-r_1}$. $\mathfrak{R}_{p-p_1}$, $\mathfrak{R}_{q-q_1}$, and $\mathfrak{R}_{r-r_1}$, which have no parallel vectors, will be called the irreducible components of $\mathfrak{R}_p$, $\mathfrak{R}_q$, and R_r. The dimension of an irreducible component is not less than two.

A parallel field of vector-spaces is called simple if it contains no parallel field of proper subspaces. In a space R_r, let V_p be a simple parallel field of vector-spaces. By [6.2], $R_r = \mathfrak{R}_p \times \mathfrak{R}_q$ $(q = r - p)$. $\mathfrak{R}_p$ cannot be factored into a product of two of its subspaces, for V_p would then not be simple; we call $\mathfrak{R}_p$ simple. To prove [6.3] we need

LEMMA 6.31: *If a Riemann space R_r has a simple parallel field of vector-spaces V_p, any parallel field of vector-spaces either contains V_p or lies in the field V_q $(q = r - p)$ of spaces of vectors normal to V_p.*

Let ξ^α_ρ $(\rho = 1, \cdots, s < r)$ span a parallel field of vector-spaces W_s which does not contain V_p. W_s cannot intersect V_p, for the intersection would be a parallel field of proper subspaces of V_p, contradicting the assumption that V_p is simple. Hence the equations $s^\rho \xi^{\alpha^2}_\rho = 0$ can have no solutions s^ρ except 0, and therefore $s \leqq q$, and $\| \xi^{\alpha^2}_\rho \|$ has rank s. Either W_s lies in V_q, or, since V_p is simple, the projection of W_s on V_p must be of dimension p, so that $s \geqq p$. If $q < p$, this contradiction proves the lemma. If $q \geqq p$, then $\| \xi^{\alpha^1}_\rho \|$ has rank p, and there is a ρ for which $\xi^1_\rho \neq 0$.

A parallel field of spaces Z_s will be constructed which intersects W_s in a parallel field of spaces whose projection on V_p, by virtue of [6.1], will be a parallel field of subspaces V_{p-1}; this contradiction will prove the lemma. Let c be a constant differing from zero and from one, let $\zeta^1_\rho = c\xi^1_\rho$ and $\zeta^\alpha_\rho = \xi^\alpha_\rho$ $(\alpha > 1)$, and let Z_s denote the field of spaces spanned by ζ^α_ρ. It is shown by Mayer (loc. cit.) that the field of spaces spanned by ξ^α_ρ is parallel if and only if functions $\gamma^\sigma_{\rho\beta}$ exist for which $\xi^\alpha_{\rho,\beta} = \gamma^\sigma_{\rho\beta}\xi^\alpha_\sigma$. Then $\zeta^1_{\rho,\beta} = c\xi^1_{\rho,\beta} = c\gamma^\sigma_{\rho\beta}\xi^1_\sigma = \gamma^\sigma_{\rho\beta}\zeta^1_\sigma$, and, when $\alpha > 1$, $\zeta^\alpha_{\rho,\beta} = \xi^\alpha_{\rho,\beta} = \gamma^\sigma_{\rho\beta}\xi^\alpha_\sigma = \gamma^\sigma_{\rho\beta}\zeta^\alpha_\sigma$, or, for all α, $\zeta^\alpha_{\rho,\beta} = \gamma^\sigma_{\rho\beta}\zeta^\alpha_\sigma$; Z_s is thus a parallel field of spaces. Since $\| \zeta^{\alpha^2}_\rho \| \equiv \| \xi^{\alpha^2}_\rho \|$ has rank s, and $s \leqq q$, the system $t^\rho \zeta^{\alpha^2}_\rho = \xi^{\alpha^2} = s^\rho \xi^{\alpha^2}_\rho$, for given s^ρ, has unique solutions $t^\rho = s^\rho$, and, if $\xi^1 \neq 0$, then $\xi^1 \neq t^\rho \zeta^1_\rho = cs^\rho \xi^1_\rho = c\xi^1$, so that ξ^α is not in Z_s; similarly, W_s does not contain Z_s. W_s and Z_s intersect in the parallel field of subspaces given by $\xi^1 = 0 = \zeta^1$, whose projection on V_p is the parallel field of proper subspaces V_{p-1} given by $\xi^1 = 0$. This contradiction proves the lemma.

It is now possible to prove

THEOREM 6.3: *A Riemann space R_r is a direct product of simple factor-spaces. Two decompositions of R_r into simple factor-spaces differ at most in the order of the factors.*

Let R_r be a Riemann space. Its decomposition into flat and irreducible components (cf. Corollary 6.21 ff.) is evidently unique, so no generality is lost in assuming that R_r is irreducible.

Either R_r is simple, or the field V_r of its tangent vector-spaces has a simple parallel field V_{p_1} of proper vector subspaces. Then $R_r = \mathfrak{R}_{p_1} \times \mathfrak{R}_{q_1}$ ($q_1 = r - p_1$), and $\mathfrak{R}_{p_1}$ is simple. Either $\mathfrak{R}_{q_1}$ is simple, or $\mathfrak{R}_{q_1} = \mathfrak{R}_{p_2} \times \mathfrak{R}_{q_2}$ ($q_2 = q_1 - p_2$), where $\mathfrak{R}_{p_2}$ is simple, and so on. At each step, the dimension of the unfactored component $\mathfrak{R}_q$ is lowered by at least two, and we arrive finally at a decomposition of R_r into a direct product of simple factor-spaces: $R_r = \mathfrak{R}_{p_1} \times \cdots \times \mathfrak{R}_{p_k}$, with $p_1 + \cdots + p_k = r$.

Let $R_r = \mathfrak{R}'_{s_1} \times \cdots \times \mathfrak{R}'_{s_l}$, with $s_1 + \cdots + s_l = r$, be a second decomposition into simple factors. The simple parallel field V'_{s_1} of spaces tangent to $\mathfrak{R}'_{s_1}$ contains the simple parallel field V_{p_1} of spaces tangent to $\mathfrak{R}_{p_1}$, or, by the lemma, it lies in V_{q_1}. If V'_{s_1} contains V_{p_1}, they coincide, for otherwise V'_{s_1} would have the parallel field of proper subspaces V_{p_1}, and would therefore not be simple; in this case $\mathfrak{R}'_{s_1} = \mathfrak{R}_{p_1}$. If V'_{s_1} lies in V_{q_1}, then V'_{s_1} either contains V_{p_2} (and, as before, coincides with it), or it lies in V_{q_2}. Proceeding in this way, we either identify $\mathfrak{R}'_{s_1}$ with one of the $\mathfrak{R}_p$, or find that V'_{s_1} lies in V_{p_k}. If V'_{s_1} and V_{p_k} did not then coincide, V_{p_k} would not be simple. In any case, V'_{s_1} coincides with one of the V_p, and $\mathfrak{R}'_{s_1}$ coincides with one of the $\mathfrak{R}_p$, which we interchange with $\mathfrak{R}_{p_1}$, and write the second decomposition as $R_r = \mathfrak{R}_{p_1} \times \mathfrak{R}'_{s_2} \times \cdots \times \mathfrak{R}'_{s_l}$. The space V'_{q_1} of vectors tangent to $\mathfrak{R}'_{q_1} = \mathfrak{R}'_{s_2} \times \cdots \times \mathfrak{R}'_{s_l}$ consists of the vectors normal to V_{p_1} and therefore coincides with V_{q_1}. Also, $p_2 + \cdots + p_k = q_1 = s_2 + \cdots + s_l$. We now treat the spaces $\mathfrak{R}_{q_1} = \mathfrak{R}'_{q_1}$ similarly, identifying $\mathfrak{R}'_{s_2}$ with a factor of $\mathfrak{R}_{q_1}$, rearranging the factors $\mathfrak{R}_p$, if necessary, so that $\mathfrak{R}'_{s_2}$ coincides with $\mathfrak{R}_{p_2}$, and finally have $\mathfrak{R}_{q_2} = \mathfrak{R}'_{q_2}$, with $p_3 + \cdots + p_k = q_2 = s_3 + \cdots + s_l$. If k were to exceed l, a stage would be reached, after $l - 1$ steps, when $\mathfrak{R}'_{s_l} = \mathfrak{R}'_{q_{l-1}} = \mathfrak{R}_{q_{l-1}} = \mathfrak{R}_{p_l} \times \cdots \times \mathfrak{R}_{p_k}$. But then the simple parallel field of spaces V'_{s_l} would contain the parallel fields of proper subspaces $V_{p_l}, \cdots, V_{p_k}$, contradicting the assumption that $\mathfrak{R}'_{s_l}$ is simple. Hence $k \leqq l$. Similarly, $l \leqq k$. Hence $k = l$, the process terminates automatically, and $\mathfrak{R}_{p_k} = \mathfrak{R}'_{p_k}$. The proof of [6.3] is now complete, for the first factorization has been so rearranged as to be identical with the second.

In terms of vector-spaces, [6.3] may be stated as

COROLLARY 6.31: *Every parallel field of vector-spaces is a product of simple parallel fields of vector-spaces. Two decompositions differ at most in the order of the factors.*

We call two vector-spaces normal if each vector of either is normal to every vector of the other. The lemma may be strengthened as

COROLLARY 6.32: *If two parallel fields of vector-spaces do not intersect, they are normal.*

7. Motions

It is seen immediately from Killing's equations $\xi_{\alpha,\beta} + \xi_{\beta,\alpha} = 0$ that the product of two motions is a motion, and that, if a product-vector is a motion, then each factor is a motion in its factor-space. If $\mathfrak{R}_p$, $\mathfrak{R}_q$, and R_r have groups of motions G_{π_1}, G_{π_2}, and G_{π_3}, it follows that $\pi_3 \geqq \pi_1 + \pi_2$.

This lower bound can be refined by using the normal form mentioned in connection with [6.21]. If $\mathfrak{R}_{p-p_1}$ has a $G_{\pi_1'}$, then $\pi_1 \geqq \pi_1' + \frac{1}{2}p_1(p_1 + 1)$; similarly for $\mathfrak{R}_{q-q_1}$. Hence $\mathfrak{R}_{r-r_1}$ has a $G_{\pi_3'}$, with $\pi_3' \geqq \pi_1' + \pi_2'$, and $R_r = \mathfrak{E}_{r_1} \times \mathfrak{R}_{r-r_1}$ has G_{π_3}, with

$$\begin{aligned}\pi_3 &\geqq \pi_3' + \frac{r_1(r_1 + 1)}{2}\\ &\geqq \pi_1' + \pi_2' + \frac{r_1(r_1 + 1)}{2}\\ &= \pi_1' + \pi_2' + \frac{p_1(p_1 + 1)}{2} + \frac{q_1(q_1 + 1)}{2} + p_1q_1.\end{aligned}$$

In this sum the term p_1q_1 counts up those rotations of $\mathfrak{E}_{r_1}$ which are not merely products of rotations of $\mathfrak{E}_{p_1}$ and $\mathfrak{E}_{q_1}$.

It will be shown that the equations hold in each case, i.e., that

THEOREM 7.1: $\pi_3 = \pi_1 + \pi_2 + p_1q_1$.

If ξ^α is a vector-field, the infinitesimal transformation $\bar{x}^\alpha = x^\alpha + \xi^\alpha\delta t$ will be a motion if and only if

$$\xi_{\alpha,\beta} + \xi_{\beta,\alpha} = 0. \tag{7.1}$$

For any vector-field,

$$\xi_{\alpha,\beta\gamma} - \xi_{\alpha,\gamma\beta} = \xi_\delta R^\delta_{\alpha\beta\gamma}. \tag{7.2}$$

From (7.1), (7.2), and the cyclic identity for R, it follows that

$$\xi_{\alpha,\beta\gamma} = \xi_\delta R^\delta_{\gamma\beta\alpha}. \tag{7.3}$$

For ξ^α to be a motion, $\xi_{\alpha,\beta}$ must satisfy this differential equation.[9]

(7.3) shows that $\xi_{\alpha,\beta\gamma}$ is breakable. By [2.2], it is a product-tensor. Hence the first (second) members of $\xi_{\alpha,\beta}$ depend only on variables of the first (second) kind, but it is not necessarily breakable. If $\xi_{\alpha,\beta}$ is also breakable, ξ_α is a product-vector and therefore the product of two motions.

If $\xi_{\alpha,\beta}$ is not breakable, its mixed members satisfy equations (7.3), of which

$$\frac{\partial \xi_{\alpha^1,\alpha^2}}{\partial x^{\beta^1}} + \xi_{\gamma^1,\alpha^2}\Gamma^{\gamma^1}_{\alpha^1\beta^1} = 0 \tag{7.4}$$

[9] Cf. C.G. §53.

is an example. Now, for a fixed value of α^2, and for a change of code, the quantities ξ_{α^1,α^2} are components of a covariant vector on each R_p. (7.4) shows that this vector is a parallel field in each R_p. Similarly, the vector ξ_{α^2,α^1} (α^1 fixed) is simultaneously a parallel field in each R_q. Further, if either of these fields is a zero field, it follows from (7.1) that the other also vanishes and ξ_α is the product of two motions. If either $\mathfrak{R}_p$ or $\mathfrak{R}_q$ fails to have parallel vectors, it then follows that there are no motions of R_r except product-motions. In this case G_{π_3} is the direct product of G_{π_1} and G_{π_2}, and $\pi_3 = \pi_1 + \pi_2$.

Now let $\mathfrak{R}_p$, $\mathfrak{R}_q$, and R_r have parallel vectors and be given in normal form. Since $\mathfrak{R}_{p-p_1}$ has no parallel vectors, $G_{\pi_1} = G_{\pi_1'} \times G_{\frac{1}{2}p_1(p_1+1)}$, and $\pi_1 = \pi_1' + \frac{1}{2}p_1(p_1 + 1)$; similarly for $\mathfrak{R}_q$. Similarly, R_r has a G_{π_3}, with $\pi_3 = \pi_3' + \frac{1}{2}r_1(r_1 + 1)$, and $\pi_3' = \pi_1' + \pi_2'$, so that $\pi_3 = \frac{1}{2}r_1(r_1 + 1) + \pi_1' + \pi_2' = \pi_1 + \pi_2 + p_1q_1$. q.e.d.

The argument also proves that

THEOREM 7.2: *G_{π_3} is the direct product of a $G_{\frac{1}{2}r_1(r_1+1)}$ and a $G_{\pi_3'}$. The group $G_{\pi_3'}$ of the irreducible component $\mathfrak{R}_{r-r_1} = \mathfrak{R}_{p-p_1} \times \mathfrak{R}_{q-q_1}$ is the direct product of $G_{\pi_1'}$ and $G_{\pi_2'}$.*

It is thus a simple matter to write down the symbols and constants of composition of G_r when those of $G_{\pi_1'}$ and $G_{\pi_2'}$ are known.

A motion is a translation if the trajectory of each point is a geodesic. Since the translations of a flat space offer no difficulty, we deal with the irreducible component $\mathfrak{R}_{r-r_1} = \mathfrak{R}_{p-p_1} \times \mathfrak{R}_{q-q_1}$ of our product-space.

If ξ^a is a translation of $\mathfrak{R}_{p-p_1}$ along geodesics C_1, and ξ^i is a translation of $\mathfrak{R}_{q-q_1}$ along geodesics C_2, and $\xi_1^\alpha = (\xi^{\alpha^1}, 0)$ and $\xi_2^\alpha = (0, \xi^{\alpha^2})$, then $\xi^\alpha = \lambda\xi_1^\alpha + \mu\xi_2^\alpha$ is a motion of $\mathfrak{R}_{r-r_1}$ whose trajectories C have projections which are the geodesics C_1 and C_2. It follows from [4.2] that the trajectories C are geodesics and ξ^α is a translation of $\mathfrak{R}_{r-r_1}$. Conversely, if ξ^α is a translation, it is the product of motions ξ^a of $\mathfrak{R}_{p-p_1}$ and ξ^i of $\mathfrak{R}_{q-q_1}$ whose trajectories C_1 and C_2 are the projections of the trajectories C of ξ^α, which are geodesics. By [4.2], C_1 and C_2 are geodesics, and ξ^a and ξ^i are translations. This proves

THEOREM 7.3: *A motion of $\mathfrak{R}_{r-r_1}$ is a translation if and only if it is the product of a translation of $\mathfrak{R}_{p-p_1}$ and a translation of $\mathfrak{R}_{q-q_1}$.*

An immediate consequence is

COROLLARY 7.31: *$\mathfrak{R}_{r-r_1}$ has a G_{ρ_3} of translations if and only if $\mathfrak{R}_{p-p_1}$ has a G_{ρ_1} of translations, and $\mathfrak{R}_{q-q_1}$ has a G_{ρ_2} of translations, and G_{ρ_3} is the direct product of G_{ρ_1} and G_{ρ_2}.*

III. The Direct Affine Product

8. The Direct Affine Product

We return to our product-manifold $\{R\} = \{P\} \times \{Q\}$ (§1), and suppose now that $\{P\}$ and $\{Q\}$ are affinely connected manifolds $\mathfrak{A}_p$ and $\mathfrak{A}_q$.[10]

---

[10] Cf. C.G. §48.

According to the superscripts of their indices, the components $A^{\alpha}_{\beta\gamma}$ of an affine connection in $\{R\}$ fall into classes which will be called *members.* An inspection of the law of transformation

$$\bar{A}^{\alpha}_{\beta\gamma}\frac{\partial x^{\delta}}{\partial \bar{x}^{\alpha}} = A^{\delta}_{\epsilon\eta}\frac{\partial x^{\epsilon}}{\partial \bar{x}^{\beta}}\frac{\partial x^{\eta}}{\partial \bar{x}^{\gamma}} + \frac{\partial^2 x^{\delta}}{\partial \bar{x}^{\beta}\partial \bar{x}^{\gamma}} \tag{8.1}$$

of $A^{\alpha}_{\beta\gamma}$ for a change of code shows (as was seen for tensors in §2) that each member behaves as an affine connection; we say that $A^{\alpha}_{\beta\gamma}$ is compound. For a mixed member, in addition, the second derivative in (8.1) vanishes for a change of code, so that each mixed member is a tensor under a change of code. A connection will be called a *product-connection* if it is breakable in a code and if its first and second members depend, respectively, in any code, only on variables of the first and second kind.

Let $\mathfrak{A}_p$ and $\mathfrak{A}_q$ have connections Λ^{a}_{bc} and Λ^{i}_{jk}. In $\{R\}$ we define $\Lambda^{\alpha^1}_{\beta^1\gamma^1} = \Lambda^{a}_{bc}$, $\Lambda^{\alpha^2}_{\beta^2\gamma^2} = \Lambda^{i}_{jk}$, and $\Lambda^{\alpha}_{\beta\gamma} = 0$ if two of its indices are of different kinds. We call $\{R\}$ with the product-connection $\Lambda^{\alpha}_{\beta\gamma}$ the *direct affine product* of $\mathfrak{A}_p$ and $\mathfrak{A}_q$ and denote it by A_r. Many of the theorems for the direct metric product depend only on the product-character of the affine connection, and can be extended easily to the direct affine product.

The paths $x^{\alpha}(t)$ of A_r are solutions of the equations $\dot{x}^{\alpha}(\ddot{x}^{\beta} + \Gamma^{\beta}_{\gamma\delta}\dot{x}^{\gamma}\dot{x}^{\delta}) - \dot{x}^{\beta}(\ddot{x}^{\alpha} + \Gamma^{\alpha}_{\gamma\delta}\dot{x}^{\gamma}\dot{x}^{\delta}) = 0$, where $\Gamma^{\beta}_{\gamma\delta} = \frac{1}{2}(\Lambda^{\beta}_{\gamma\delta} + \Lambda^{\beta}_{\delta\gamma})$ is the symmetric part of $\Lambda^{\alpha}_{\beta\gamma}$. The proof of [4.2] may be repeated exactly, with "geodesic" replaced by "path," and "arc-lengths" by "affine parameters."

The definitions of "plane at a point" and "plane" are valid if "geodesic" is replaced by "path." Evidently the subspaces A_p and A_q are planes in A_r.

An *affine normal coördinate system y at a point P* is one in which the path through P in the direction ξ^{α} has equations $y^{\alpha} = \xi^{\alpha}t$. With this understanding, [4.3] holds for affine normal coördinate systems. The proofs of [4.4] – [4.51] can then be repeated exactly with "geodesic" replaced by "path."

Parallel displacement with respect to $\Lambda^{\alpha}_{\beta\gamma}$ between neighboring points or along a curve has the product-character expressed in [5.1], and [5.2] therefore holds for A_r. These theorems hold also for displacements parallel with respect to $\Gamma^{\alpha}_{\beta\gamma}$. The study of curvature leading to [5.3] fails for want of a metric tensor.

Theorem 6.1 and its corollaries carry over without modification. Mayer's proof of [6.2] for an $\mathfrak{R}_p$ depended on the fact that the family of spaces V_{p-p_1} of vectors normal to a V_{p_1} is a parallel field of vector-spaces if V_{p_1} is. In an $\mathfrak{A}_p$ there is no such unique vector-space complementary to a V_{p_1}, and this method of proof of [6.2] is not valid. No direct proofs of extensions of the remaining results of §6 have been found.

The infinitesimal transformation $\bar{x}^{\alpha} = x^{\alpha} + \xi^{\alpha}\delta t$ will be a *collineation* of A_r (will carry paths into paths) if and only if[11]

$$\begin{cases} \xi^{\alpha}_{;\beta\gamma} = \xi^{\delta}B^{\alpha}_{\beta\gamma\delta}, \\ \Omega^{\alpha}_{\beta\gamma;\delta}\xi^{\delta} + \Omega^{\alpha}_{\delta\gamma}\xi^{\delta}_{;\beta} + \Omega^{\alpha}_{\beta\delta}\xi^{\delta}_{;\gamma} - \Omega^{\delta}_{\beta\gamma}\xi^{\alpha}_{;\delta} = 0, \end{cases} \tag{8.2}$$

[11] Cf. C.G. §58.

where $\Omega^{\alpha}_{\beta\gamma} = \frac{1}{2}(\Lambda^{\alpha}_{\beta\gamma} - \Lambda^{\alpha}_{\gamma\beta})$, the semi-colon indicates covariant differentiation with respect to $\Gamma^{\alpha}_{\beta\gamma}$, and $B^{\alpha}_{\beta\gamma\delta}$ is defined in terms of $\Gamma^{\alpha}_{\beta\gamma}$ precisely as $R^{\alpha}_{\beta\gamma\delta}$ was in terms of the Christoffel symbols $\Gamma^{\alpha}_{\beta\gamma}$. By virtue of the product-character of $\Lambda^{\alpha}_{\beta\gamma}$ (and hence of $\Gamma^{\alpha}_{\beta\gamma}$, $\Omega^{\alpha}_{\beta\gamma}$, and $B^{\alpha}_{\beta\gamma\delta}$), it follows from (8.2) that the product of a collineation of $\mathfrak{A}_p$ and a collineation of $\mathfrak{A}_q$ is a collineation of A_r.

We prove a weakened extension of Theorem 7.2.

THEOREM 8.1: *If either $\mathfrak{A}_p$ or $\mathfrak{A}_q$ fails to have a parallel vector-field, the group of collineations of A_r is the direct product of the groups of collineations of $\mathfrak{A}_p$ and $\mathfrak{A}_q$.*

If ξ^{α} is a collineation of A_r which is not a product of collineations of $\mathfrak{A}_p$ and $\mathfrak{A}_q$, it follows from the first of (8.2), as in the proof of [7.2], that $\xi^{\alpha^1}_{;\beta^2}$ is a parallel field in A_p (β^2 fixed) and in A_q (α^1 fixed); similarly for $\xi^{\alpha^2}_{;\beta^1}$. If, then, either $\mathfrak{A}_p$ or $\mathfrak{A}_q$ fails to have a parallel field, each of these fields vanishes. This means that $\xi^{\alpha^1}_{;\beta^2} = \partial\xi^{\alpha^1}/\partial x^{\beta^2} = 0$ and $\xi^{\alpha^2}_{;\beta^1} = \partial\xi^{\alpha^2}/\partial x^{\beta^1} = 0$, so that ξ^{α} is a product of two collineations, and the group of A_r is the direct product of the groups of $\mathfrak{A}_p$ and $\mathfrak{A}_q$. q.e.d.

The more exact results of [7.1] and [7.2] cannot be extended without more definite information on an $\mathfrak{A}_p$ with parallel fields.

A collineation may be called a *translation* if the trajectory of each point is a path. [7.3] and [7.31] depend only on [4.2]. Using the extension of [4.2] to an A_r and replacing "motion" by "collineation", the proofs of [7.3] and [7.31] may be repeated exactly for the case when one of the factor-spaces has no parallel fields.

IV. REMARKS ON GENERAL PRODUCT-SPACES

$R_r = \mathfrak{R}_p \times \mathfrak{R}_q$ is a general Riemann product-space if $g_{\alpha^1\alpha^2} \neq 0$. A two-space which is a general product will have $ds^2 = du^2 + dv^2 + 2F(u, v)\, dudv$ with $F^2 < 1$. It is trivially easy to show that, if $\bar{R}_r$ is in one-one isometric or conformal correspondence with a product-space R_r, then $\bar{R}_r$ is a product-space which is direct if and only if R_r is.

9. General Product-spaces

The peculiar simplicity of the direct product followed from the product-character of $g_{\alpha\beta}$, $\Gamma^{\alpha}_{\beta\gamma}$, etc., and the simple behaviour of product-tensors. In attempting to apply the methods used in studying the direct product, it is natural to try to restrict $g_{\alpha^1\alpha^2}$ so as partially to preserve the simplicity of the direct product. Each such effort has been found to lead back to the direct product. Since these results are mainly negative, they will be summarized without proof.

The complication of $g_{\alpha\beta}$ makes itself felt at a very elementary level. In an ordinary Riemann space $\mathfrak{R}_p$, it is usual to speak of ξ^a and associated functions $\xi_a = g_{ab}\xi^b$ as the contravariant and covariant components of the same vector ξ. If ξ and η are vectors in $\mathfrak{R}_p$ and $\mathfrak{R}_q$, we define their contravariant product ζ^{α}_{i} to have components $\zeta^{\alpha}_{i} = (\xi^{\alpha^1}, \eta^{\alpha^2})$, and their covariant product $\zeta_{v\alpha}$ to have components $\zeta_{v\alpha} = (\xi_{\alpha^1}, \eta_{\alpha^2})$. It is easy to show that $\zeta_{v\alpha} = g_{\alpha\beta}\zeta^{\beta}_{i}$, that is, the covariant and contravariant products of two factor-vectors coincide in R_r, if and only

if R_r is the direct product of $\mathfrak{R}_p$ and $\mathfrak{R}_q$. Further, the contravariant (covariant) product of a given pair of vectors will necessarily be also the covariant (contravariant) product of some pair of vectors if and only if R_r is direct.

The length of the elementary vector $\zeta_t^\alpha = (\xi^{\alpha^1}, 0)$ calculated in R_r obviously agrees with its length calculated in $\mathfrak{R}_p$. It can be shown that $\zeta_{v\alpha} = (\xi_{\alpha^1}, 0)$ has the same length in R_r as in $\mathfrak{R}_p$ if and only if R_r is direct. The contravariant product of two unit vectors has squared length $|\zeta_t|^2 = 2[1 + \cos(\xi, \eta)]$. Their covariant product ζ_v will obey a similar law if and only if the lengths of elementary covariant vectors are preserved, i.e., R_r is direct.

The difficulties encountered with covariant product-vectors may be removed in exchange for similar difficulties with contravariant product-vectors in the following rather artificial way. Instead of using g_{ab} and g_{ij} and defining further functions $g_{\alpha^1\alpha^2}$, we use g^{ab} and g^{ij}, denote them by $h^{\alpha^1\beta^1}$ and $h^{\alpha^2\beta^2}$ and define further functions $h^{\alpha^1\beta^2} \neq 0$ in such a way that $h^{\alpha\beta}$ is symmetric and positive definite. In this way the manifold $\{R\}$ is again made into a Riemann space which may be called a general contravariant product R^r of $\mathfrak{R}_p$ and $\mathfrak{R}_q$. The remarks above on R_r may then be paraphrased in an obvious way for R^r. Further, we may define an $(R_r)^r$ associated with R_r by taking $h^{\alpha^1\beta^2}$ to be the functions $g^{\alpha^1\alpha^2}$ occurring in the inverse $g^{\alpha\beta}$ of $g_{\alpha\beta}$ in R_r; similarly, a given R^r has associated with it an $(R^r)_r$ constructed by replacing $h_{\alpha^1\beta^1}$ and $h_{\alpha^2\beta^2}$ in the inverse $h_{\alpha\beta}$ of $h^{\alpha\beta}$ for R^r by $g_{\alpha^1\beta^1} = g_{ab}$ and $g_{\alpha^2\beta^2} = g_{ij}$. It is easily shown that these definitions are legitimate in the sense that if the metric of R_r (or of R^r) is positive definite, then so is the metric of $(R_r)^r$ (or of $(R^r)_r$). It can be proved that $h^{\alpha\beta}$ for $(R_r)^r$ equals $g^{\alpha\beta}$ for R_r if and only if R_r is direct; similarly for R^r and $(R^r)_r$. Finally, $((R_r)^r)_r = R_r$ if and only if R_r is direct.

In the direct product, it was advantageous to know that the product of two motions or of two parallel vectors was a motion or was parallel. When the necessary conditions for the truth of these statements in a general R_r are required to be satisfied by arbitrary product-vectors, it is seen that the contravariant or (independently) the covariant product of two motions or of two parallel vectors is necessarily a motion or a parallel vector if and only if R_r is direct. This completes our remarks on vectors.

In a general product-space, under a change of code, tensors are compound. $\Gamma^\alpha_{\beta\gamma}$ is compound, and its mixed members are tensors. The first (second) members of $[\beta\gamma, \alpha]$ depend only on variables of the first (second) kind, but $\Gamma^\alpha_{\beta\gamma}$ does not necessarily have this property. The subspaces R_p and R_q need not be totally geodesic in R_r. The projections of a geodesic are not necessarily geodesics of the factor-spaces, so that the product of two normal coördinate systems is not necessarily a normal code. Our earlier results are thus incapable of immediate extension.

By examining in detail the equations for parallelism along a curve, separate interpretations can be read off for each of the various members of $\Gamma^\alpha_{\beta\gamma}$. By using §6, it can be shown that if $g_{\alpha^1\alpha^2}$ are so defined in R_r that $\Gamma^\alpha_{\beta\gamma}$ is breakable and $\Gamma^{\alpha^1}_{\beta^1\gamma^1} = \Gamma^a_{bc}$ and $\Gamma^{\alpha^2}_{\beta^2\gamma^2} = \Gamma^i_{jk}$, then R_r is direct.

In spite of the awkwardness of general product-spaces, sufficient conditions can be given that an arbitrary Riemann space R_r be a general product-space. Suppose that a coördinate system exists in which the coördinates can be so grouped that, when transformations of each group separately are alone permitted, $\Gamma^{\alpha}_{\beta^1 \gamma^2} = 0$; then R_r is a general product-space and this coördinate-system is a code. Finally, if R_r is a Riemann space with a transitive group of motions G_r which is the direct product of a G_{π_1} and a G_{π_2}, then R_r is a general product of an $\mathfrak{R}_{r_1}$ and an $\mathfrak{R}_{r_2}$, r_1 and r_2 being the ranks of the matrices of the symbols of G_{π_1} and G_{π_2}.

PRINCETON UNIVERSITY.

Annals of Mathematics
Vol. 40, No. 4, October, 1939

NON-ALTERNATING INTERIOR RETRACTING TRANSFORMATIONS

By G. T. Whyburn

(Received April 24, 1939)

1. Introduction

It is well known that any topological n-cell, and hence certainly any simple arc, is a retract of any containing topological space; and any topological polyhedron, in particular any simple closed curve, is a retract of some of its suitably chosen neighborhoods in any containing space. When anything more than continuity is required of the "retracting" transformation, however, the situation is obviously markedly altered and yet it seems not to have been the subject of much investigation.

In this paper conditions will be developed under which a locally connected continuum M can be retracted onto an arc or a simple closed curve by transformations of the type indicated in the title. The principal results are: (1) for M to be retractable into an arbitrarily given arc axb by such a transformation it is necessary and sufficient that M be identical with the cyclic chain $C(a, b)$ in M; and (2) for M to be retractable into a given simple closed curve J by such a transformation it is necessary and sufficient that M be cyclic and not unicoherent about J. In the proofs, no use will be made of the facts concerning retracts mentioned above.

It will be recalled that a continuous transformation $f(S) = R$ is said to be *retracting*[1] (and R is then a *retract* of S) provided $R \subset S$ and for each $x \in R$, $f(x) = x$. Also, a continuous transformation $T(A) = B$ is said to be *interior*[2] provided sets open in A map into sets open in B, and T is said to be *non-alternating*[3] provided that for no two points $x, y \in B$ does $T^{-1}(x)$ separate two points of $T^{-1}(y)$ in A.

Throughout the paper M will denote a compact locally connected continuum. For any two points a and b of M, the *cyclic chain* $C(a, b)$ in M is the set consisting of a, b, the set K of all points separating a and b in M, and all cyclic elements of M containing two points of $K + a + b$. More simply, $C(a, b)$ consists of all simple arcs in M of the form axb.

2. Separation and subdivision

(2.1) Lemma. *Let $M = C(a, b)$, let A and B be disjoint continua in M containing a and b respectively and let $a'xb'$ be any arc in M with $a'xb' \cdot A = a'$,*

---

[1] See Borsuk, *Fundamenta Mathematicae*, vol. 17(1931) p. 152.
[2] Stoilow, *Annales Scientifiques de l'Ecole Normale Supérieure*, vol. 63(1928) pp. 347–382.
[3] See my paper in *American Journal of Mathematics*, vol. 56(1934), pp. 294–302.

$a'xb' \cdot B = b'$. There exists a set X with $X \cdot a'xb' = x$ which irreducibly separates M between A and B into just two components.

PROOF. If x separates A and B in M, we have only to set $X = x$. If not, then $M - x$ is connected and locally connected. Hence by a result previously established[4] we can find a decomposition

$$M - x = R_a + F + R_b,$$

where R_a and R_b are disjoint connected open sets containing $A + a'x - x$ and $B + xb' - x$ respectively and where $F = F(R_a) = F(R_b)$ (boundaries taken relative to $M - x$). Thus if we set $F + x = X$ we have

$$M - X = R_a + R_b,$$

where $F(R_a) = F(R_b) = X$ (boundaries taken relative to M), and our lemma is satisfied.

Let the locally connected continuum M be a cyclic chain $C(a, b)$. By a *subdivision*[5] σ of M will be meant a finite, linearly ordered set of disjoint irreducible cuttings of M between a and b

$$a = X_0, X_1, X_2, X_3, \cdots, X_n, X_{n+1} = b,$$

where X_i $(1 \leqq i \leqq n)$ cuts M into two connected sets $M_a(X_i)$, $M_b(X_i)$ so that $M_a(X_i) \supset a + \sum_1^{i-1} X_j$ and $M_b(X_i) \supset b + \sum_{i+1}^{n} X_j$. The set

$$X_{i-1} + X_i + M_a(X_i) \cdot M_b(X_{i-1}) = I_i$$

will be called the *internal* from X_{i-1} to X_i $(a = X_0, b = X_{n+1})$. A set of the form $M_a(X_i) \cdot M_b(X_{i-1})$ will be called an *open interval* of σ. A subdivision σ' will be called a *refinement* of σ provided it is obtained from σ by inserting additional elements. (The sets X_i are called *elements* of the subdivision.)

Still assuming $M = C(a, b)$, let axb be any arc in M from a to b. We proceed to prove

(2.2) LEMMA. *Given any subdivision σ_0 of M each element of which contains just one point of axb and any $\epsilon > 0$, there exists a refinement σ of σ_0 such that if X_i is any element of σ, then $axb \cdot X_i$ is just one point and*[6] $V_\epsilon(X_i) \supset I_i + I_{i+1}$.

PROOF.[5] For convenience we will suppose the metric in M so chosen[7] that all sets of the form $V_r(x)$, $x \in M$, are connected. Let $e = \epsilon/5$ and let us cover M with a finite number of the sets $V_e(x)$, say $V_1 = V_e(x_1)$, $V_2 = V_e(x_2)$, $\cdots$, $V_m = V_e(x_m)$.

Now consider V_1. If $\bar{V}_1 \supset b$ we need go no further. Hence we suppose $\bar{V}_1 \cdot b = 0$ and let X_{i+1} be the first element of σ_0 such that $M_a(X_{i+1}) \supset \bar{V}_1$.

---

[4] See my paper in *Bulletin of the American Mathematical Society*, vol. 37(1931) p. 734.

[5] Compare with my paper in the *Duke Mathematical Journal*, vol. 5(1939), pp. 647–655, where results very closely related to (2.2) and (3.1) are established.

[6] $V_r(X)$ denotes the r-neighborhood of X, i.e., the set of all points x at a distance $< r$ from X.

[7] See Mazurkiewicz, *Fundamenta Mathematicae*, vol. 1(1920), p. 27.

Now if $\overline{V_{2e}(x_1)} \cdot X_{i+1} \neq 0$ we need go no further. If this is not so, we proceed as follows. Setting $\alpha = M_a(X_i) + X_i$, $\beta = M_b(X_{i+1}) + X_{i+1}$ and letting x' be a point of V_1 in the open interval (X_i, X_{i+1}) of σ_0, we can find[8] a continuum $\alpha x'\beta$ in M which is a "simple arc" between the continua α and β, i.e., if α and β are shrunk to points we have an ordinary simple arc going through x'.

We shall treat first the case where $axb \cdot V_{2e}(x_1) = 0$. On $\alpha x'\beta$ proceeding from x' toward α and from x' toward β let α' and β' respectively denote the first "points" of the set $\alpha + \beta + axb$.

(i) If $\alpha' \neq \alpha$ and $\beta' \neq \beta$, let c be the first and d the second of the points α' and β' on axb in the order a, b and let $E =$ arc cx' of $\alpha x'\beta$, $F =$ arc $x'd$ of $\alpha x'\beta$.

(ii) If $\alpha' \neq \alpha$ and $\beta' = \beta$, let $c = \alpha'$, $d = axb \cdot X_{i+1}$, $E =$ arc cx' of $\alpha x'\beta$, and $F = x'\beta$ of $\alpha x'\beta$.

(iii) If $\alpha' = \alpha$ and $\beta' \neq \beta$, let $c = axb \cdot X_i$, $d = \beta'$, $E = \alpha x'$ of $\alpha x'\beta$ and $F = x'd$ of $\alpha x'\beta$.

(iv) If $\alpha' = \alpha$ and $\beta' = \beta$, let $c = axb \cdot X_i$, $d = axb \cdot X_{i+1}$, $E = \alpha x'$ of $\alpha x'\beta$, $F = x'\beta$ of $\alpha x'\beta$.

In any case, let y be the last point of $\overline{V_{2e}(x_1)}$ on $\alpha' x'\beta'$ in the order α', β' and set

$$A = \alpha + ac \text{ (of } axb) + E + \bar{V}_1$$

$$B = \beta + db \text{ (of } axb) + F - (x'y - y).$$

Let x be a point on axb between c and d. Since $cxd \cdot A = c$, $cxd \cdot B = d$, we may apply (2.1) and obtain a set X with $cxd \cdot X = x$ which separates M irreducibly between A and B into just two components.

Now in case $axb \cdot V_{2e}(x_1) \neq 0$, let d be the last point of $\overline{V_{2e}(x)}$ on axb in the order a, b and let x' be chosen as before. We can either join x' to $ad - d$ by an arc $px' \subset M_a(X_{i+1})$ such that $px' \cdot axb = p$ or we can join x' to X_i by an arc px' such that $p \in X_i$ and $px' \cdot axb = 0$. If $p \in axb$, set $c = p$; if not, set $c = X_i \cdot axb$. Let x be between c and d on axb. Then if we set $A = \alpha + ac + px'$, $B = \beta + db$, we can proceed as before to find the set X.

Clearly X is on the "b side" of V_1, i.e., $M_a(X) \supset V_1$; and $X \cdot V_{2e}(x_1) \neq 0$ since $A \cdot V_{2e}(x_1) \neq 0 \neq B \cdot \overline{V_{2e}(x_1)}$. In exactly the same manner we construct a set Y on the "a side" of V_1 such that $Y \cdot V_{2e}(x_1) \neq 0$. Let us add X and Y to σ_0 and call σ_1 the resulting refinement of σ_0.

Similarly for V_2 we obtain a refinement σ_2 of σ_1 which retains the properties of σ_1 and in addition contains two elements W and Z each intersecting $V_{2e}(x_2)$ and such that $V_2 \subset M_b(W) \cdot M_a(Z)$. Proceeding in this manner to V_m we finally obtain a refinement σ_m of σ_0, which we call σ, which retains the essential properties of σ_0 and in addition has the property that for any $i \leqq m$, there are two elements X_j and X_k of σ intersecting $V_{2e}(x_i)$ and such that $V_i \subset M_b(X_j) \cdot M_a(X_k)$. Since $e = \epsilon/5$ and the sets V_i cover M, clearly this is equivalent to our lemma.

[8] See (1.1) of the paper referred to in [5].

3. Retractions into arcs

(3.1) THEOREM. *In order that a locally connected continuum M be retractable into an arbitrarily given arc axb in M by a non-alternating interior transformation it is necessary and sufficient that M be identical with the cyclic chain $C(a, b)$ in M.*

PROOF. We first show that the condition is sufficient. By lemma (2.2) we can find a subdivision σ_1 of M satisfying the conclusion of this lemma for $\epsilon = 1$. Let us set $\sigma_1 = \sigma_1'$ and suppose we have constructed σ_i and σ_i' for all $i < n$. Then let us choose points $x_1, x_2, \cdots x_k$ on axb such that the diameter of each of the arcs $ax_1, x_1x_2, \cdots, x_kb$ is $< 1/n$. Then by lemma (2.2), we can find a refinement σ_n of σ_{n-1}' satisfying the conclusion of that lemma for $\epsilon = 1/n$; and by applying lemma (2.1) successively to the points $x_1, x_2 \cdots x_k$ we can obtain a refinement σ_n' of σ_n such that each of the points $[x_i]$ belongs to exactly one element of σ_n' and such that σ_n' still satisfies the conclusion of lemma (2.2) for $\epsilon = 1/n$. To see this, consider x_i. Suppose x_i lies within the interval YZ of σ_n (if x_i belongs to an element of σ_n, we need make no construction for x_i). Then $axb \cdot YZ$ is an interval cx_id of axb; and if we set $M_a(Y) + Y = A$, $M_b(Z) + Z = B$ and apply lemma (2.1), we get a set X_i with $axb \cdot X_i = x_i$ which separates M irreducibly between A and B into just two components. If we add such a set X_i to σ_n for each x_i not on an element of σ_n, we obtain the required subdivision σ_n'.

Thus we have set up an infinite monotone sequence $\sigma_1', \sigma_2', \cdots$ such that for each n, σ_n' satisfies the conditions of lemma (2.2) for $\epsilon = 1/n$ and in addition such that each interval $I = X_iX_{i+1}$ of σ_n' intersects axb in an arc of diameter $< 1/n$.

We now define a transformation $f(p)$ of M into axb as follows. If p belongs to some element X_p of σ_n' for some n, let $f(p) = axb \cdot X_p$; if not, then let $Z_p = \prod_1^\infty I_n$, where I_n is the (minimum) interval of σ_n' containing p and let $f(p) = axb \cdot Z_p$.

Since for each n, $I_n \cdot axb$ is an interval of axb of diameter $< 1/n$, it follows that in either case $f(p)$ is a single point so that f is single valued. Obviously $f(p) = p$ for $p \epsilon axb$ so that f is retracting. Furthermore, since for any point p other than a or b and any $\epsilon > 0$ we can find an open interval $E = (X_j, X_k)$ of some σ_n' containing p whose intersection with axb is an open interval (x_jx_k) of axb of diameter $< \epsilon$ and since $f(E) = (x_jx_k)$, it follows that f is continuous. Finally, since for any $x \epsilon axb - (a + b)$, it follows by the construction of the subdivisions σ_n' that $f^{-1}(x)$ separates M irreducibly between a and b into just two components R_a and R_b which map under f onto the arcs $ax - x$ and $xb - x$ respectively, and since $a = f^{-1}(a)$, $b = f^{-1}(b)$, it is readily seen that f is non-alternating and interior.

To show the condition necessary,[5] let $f(M) = axb$ be a non-alternating, interior and retracting but suppose, contrary to our statement, that there is a component R of $M - C(a, b)$. The boundary of R is a single point x. If $x = a$ (or b), then since f is interior, $f(R) \supset axb - a$ so that R contains a point

of $f^{-1}(b)$; but since R does not contain b, this contradicts the fact that f is non-alternating. If $a \neq x \neq b$, then since f is interior, $f(R)$ must contain either $ax - x$ or $xb - x$; but since R contains neither a nor b, again it follows that f alternates.

REMARK. Attention may be called to the fact that the transformation f set up in the sufficiency part of the proof is so defined that $f^{-1}(a) = a, f^{-1}(b) = b$. Also, for any $x \in axb - (a + b), f^{-1}(x)$ separates M irreducibly between a and b into just two components. It results from this that *in case M is unicoherent, f is necessarily monotone.*

Now suppose axb is a simple arc and $g(x)$ is any interior transformation defined on axb. Let M be any locally connected continuum containing axb and such that M is identical with the cyclic chain $C(a, b)$. Let f be an interior transformation retracting M into axb as given by the above theorem. Then the transformation $\varphi = gf$ is interior; and since for $x \in axb$ we have $f(x) = x$ and thus $\varphi(x) = g(x), \varphi$ is an extension of $g(x)$ to M. Hence we have

(3.2) EXTENSION THEOREM. *Any interior transformation on a simple arc axb can be extended (interiorly) to any locally connected continuum M containing axb which is of the form $M = C(a, b)$.*

Now let M be any locally connected continuum whatever and let axb be any arc in M. Since[3] M can be retracted into any A-set in M by a monotone transformation, there exists a monotone transformation $g(x)$ which retracts M into $C(a, b)$. Applying (3.1), we get a non-alternating interior transformation $f(y)$ which retracts $C(a, b)$ into axb. The transformation $\varphi(x) = fg(x)$ is then[3] non-alternating and it retracts M into axb. Since $\varphi(x) \equiv f(x)$ for $x \in C(a, b)$ and f is interior, it follows that when considered on $C(a, b)$ alone, φ is interior. Thus we have

(3.3) THEOREM. *Any locally connected continuum M can be retracted into any arc axb in M by a non-alternating transformation $\varphi(x)$ which is interior when considered on the cyclic chain $C(a, b)$ alone.*

4. Retractions into simple closed curves

A locally connected continuum M is said to be unicoherent about a simple closed curve J in M[9] provided that for every decomposition $M = H + K$ where H and K are closed sets such that $H \cdot J$ is an arc xry and $K \cdot J$ an arc xsy, the points x and y belong to the same component of $H \cdot K$.

(4.1) LEMMA. *If M is cyclic and not unicoherent about the simple closed curve $J \subset M$, there exist disjoint closed sets X and Y intersecting J in single points x and y respectively and disjoint connected regions R and S in M containing the open arcs xry and xsy respectively of J and such that $M - (X + Y) = R + S$ and $F(R) = X + Y = F(S)$.*

PROOF. By hypothesis and a theorem of W. A. Wilson[9] there exists a decomposition $M = H + K$ where H and K are continua such that $H \cdot J$ is an arc

[9] See W. A. Wilson, *American Journal of Mathematics*, vol. 55(1933), pp. 135–145.

amb and $K \cdot J$ an arc anb. Let U be the component of $M - K$ containing $amb - (a + b)$ and let $W = M - U$. It follows from our hypothesis that $F(U) = A + B$ where A and B are disjoint and closed and $a \in A, b \in B$.

Let us decompose $\bar{U}$ upper semi-continuously[10] into the sets A, B and individual points of U. Let U' be the hyperspace of this decomposition and $w(\bar{U}) = U'$ the associated continuous transformation. Clearly U' is a locally connected continuum. Furthermore, if $\alpha = w(A)$, $\beta = w(B)$, then since M is cyclic and neither A not B cuts $\bar{U}$ (so that α and β are non-cut points of U'), it follows that U' is identical with the cyclic chain $C(\alpha, \beta)$ in U'. Hence if $\alpha\mu\beta$ denotes the arc $w(amb)$, by (3.1) there exists a non-alternating interior transformation $g(x)$ which retracts U' into $\alpha\mu\beta$ so that $\alpha = g^{-1}(\alpha)$, $\beta = g^{-1}(\beta)$ and for every $\mu \in \alpha\mu\beta - (\alpha + \beta)$, $g^{-1}(\mu)$ separates U' irreducibly between α and β into just two components.

Let v be a point of $\alpha\mu\beta - (\alpha + \beta)$ and let $V = w^{-1}g^{-1}(v)$. Since U is connected and V is a compact subset of U, there exists a (compact) continuum N with $V \subset N \subset U$. Since $gw(N)$ contains neither α nor β, we can find interior points ξ and η on $\alpha\mu\beta$ so that we have the order $\alpha, \xi, v, \eta, \beta$ and so that the open segment $Q = \xi v\eta - (\xi + \eta)$ of $\alpha\mu\beta$ contains $gw(N)$. Let

$$X = w^{-1}g^{-1}(\xi), \qquad Y = w^{-1}g^{-1}(\eta)$$

$$R = w^{-1}g^{-1}(Q), \qquad S = M - (R + X + Y).$$

These sets satisfy the conditions of our lemma as will now be shown.

Obviously $X \cdot J$ and $Y \cdot J$ are points x and y respectively and $R \cdot J$ and $S \cdot J$ are the open arcs xry and xsy respectively of J. Also $F(R) = X + Y = F(S)$, since each of the sets $g^{-1}(\xi)$ and $g^{-1}(\eta)$ separates U' irreducibly between α and β into just two components and w is topological on U. For the same reason, the closure of each component of R must intersect both X and Y. Hence each such component must intersect V and thus N. As $N \subset R$ it results that R is connected. Similarly, each component of S must intersect either A or B and hence W; and since W is connected, S is connected. This completes the proof.

(4.2) Theorem. *In order that a locally connected continuum M be retractable into a given simple closed curve $J \subset M$ by a non-alternating interior transformation it is necessary and sufficient that M be cyclic and not unicoherent about J.*

We shall first show that the conditions are sufficient. Let X, Y, R and S be sets satisfying Lemma (4.1). Let us decompose M into the sets X, Y and individual points of $M - (X + Y)$. Call M' the hyperspace of this decomposition and let $h(M) = M'$ be the associated transformation. Let $h(X) = a'$, $h(Y) = b'$, $h(J) = J' = h(xry) + h(xsy) = a'r'b' + a's'b'$, $h(R) = R'$, $h(S) = S'$. Then by (4.1) it follows that $\bar{R}'$ is identical with its cyclic chain $C(a', b')$ and $\bar{S}'$ is identical with its cyclic chain $C(a', b')$. Hence by (3.1) there exists

---

[10] See R. L. Moore, *Transactions of the American Mathematical Society*, vol. 27 (1925) pp. 416–428.

non-alternating interior functions φ_r and φ_s retracting $\bar{R}'$ and $\bar{S}'$ into $a'r'b'$ and $a's'b'$ respectively so that $\varphi_r^{-1}(a') = \varphi_s^{-1}(a') = a'$, $\varphi_r^{-1}(b') = \varphi_s^{-1}(b') = b'$. Thus if we let $\varphi(x)$ equal $\varphi_r(x)$ on $\bar{R}'$ and $\varphi_s(x)$ on $\bar{S}'$, then φ is a non-alternating interior function retracting M' into J'.

We now define

$$f(x) = h^{-1}[\varphi h(x)]\cdot J.$$

(NOTE: This equals $h^{-1}\varphi h(x)$ except when $x \epsilon X + Y$.) Clearly $f(x)$ retracts M into J. Since φh is interior, if G is an open set in M, $\varphi h(G)$ is open in J'; hence $h^{-1}[\varphi h(G)]$ is open in M, so that the intersection of this set with J is open in J. Accordingly f is interior. To show that f is non-alternating, let $t \epsilon J$. Then $f^{-1}(t) = h^{-1}\varphi^{-1}h(t)$. If $h(t)$ is either a' or b', say a', then $\varphi^{-1}h(t) = a'$ and $f^{-1}(t) = X$, which does not separate M. If $a' \neq h(t) \neq b'$, then $\varphi^{-1}h(t)$ cuts either $\bar{R}'$ or $\bar{S}'$, say $\bar{R}'$, irreducibly between a' and b'. Thus any component of $M - f^{-1}(t)$ must intersect either X or Y; and since $S + X + Y$ is connected, it follows that $M - f^{-1}(t)$ is connected. Thus as no set $f^{-1}(t)$ separates M, f is non-alternating.

To prove the conditions necessary, let $f(M) = J$ be non-alternating, interior and retracting, where M is a locally connected continuum and J is a simple closed curve in M. Let $C(J)$ be the cyclic element of M containing J. Then $M = C(J)$. For if not, there exists a component R of $M - C(J)$; and if u denotes the boundary of R and v is a point of $R - R\cdot f^{-1}f(u)$, $f(u) = x$, $f(v) = y$, then $f^{-1}(x)$ would separate the points y and v of $f^{-1}(y)$ in M contrary to the non-alternating property of f. Hence M is cyclic. To see that M is not unicoherent about J, we have only to take points a and b on J decomposing J into arcs axb and ayb and set $H = f^{-1}(axb)$, $K = f^{-1}(ayb)$. This gives $M = H + K$, $H\cdot J = axb$, $K\cdot J = ayb$; and since $f(H\cdot K) = a + b$, obviously a and b lie in different components of $H\cdot K$.

It will be noted that the transformation $f(x)$ set up in the sufficiency part of the proof just given is so constructed as to have the following additional properties:

(4.21) *There exist open intervals α and β of J such that if $x \epsilon \alpha$, $y \epsilon \beta$, $f^{-1}(x) + f^{-1}(y)$ separates M irreducibly into just two components.*

(4.22) *If $p^1(M) = 1$, f is monotone (or if $p^1(M) = k$, then for each $x \epsilon J$, $f^{-1}(x)$ has at most k components).*[11]

The second of these is an immediate consequence of the first.

The same device as used above to obtain (3.2) yields

(4.3) EXTENSION THEOREM. *Any interior transformation on a simple closed curve J can be extended (interiorly) to any cyclic locally connected continuum containing J which is not unicoherent about J.*

Also the method of obtaining (3.3) from (3.1) when applied to (4.2) yields

(4.4) THEOREM. *If M is any locally connected continuum and J is a simple closed curve in M about which M is not unicoherent, M can be retracted into J by a*

---

[11] $p^1(M)$ denotes the 1-dimensional connectivity number of M.

non-alternating transformation which is interior when considered only on the cyclic element $C(J)$ of M containing J.

In conclusion we remark that since by a result of Wilson,[9] if a locally connected continuum M is multicoherent, it contains some simple closed curve about which it is not unicoherent, we have that *In order for M to be retractable into a simple closed curve by a non-alternating interior transformation it is necessary and sufficient that M be cyclic and multicoherent.*

University of Virginia.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

ON A STATIONARY SYSTEM WITH SPHERICAL SYMMETRY CONSISTING OF MANY GRAVITATING MASSES

BY ALBERT EINSTEIN

(Received May 10, 1939)

If one considers Schwarzschild's solution of the static gravitational field of spherical symmetry

$$ds^2 = -\left(1 + \frac{\mu}{2r}\right)^4 (dx_1^2 + dx_2^2 + dx_3^2) + \left(\frac{1 - \frac{\mu}{2r}}{1 + \frac{\mu}{2r}}\right)^2 dt^2 \tag{1}$$

it is noted that

$$g_{44} = \left(\frac{1 - \frac{\mu}{2r}}{1 + \frac{\mu}{2r}}\right)^2$$

vanishes for $r = \mu/2$. This means that a clock kept at this place would go at the rate zero. Further it is easy to show that both light rays and material particles take an infinitely long time (measured in "coördinate time") in order to reach the point $r = \mu/2$ when originating from a point $r > \mu/2$. In this sense the sphere $r = \mu/2$ constitutes a place where the field is singular. (μ represents the gravitating mass.)

There arises the question whether it is possible to build up a field containing such singularities with the help of actual gravitating masses, or whether such regions with vanishing g_{44} do not exist in cases which have physical reality. Schwarzschild himself investigated the gravitational field which is produced by an incompressible liquid. He found that in this case, too, there appears a region with vanishing g_{44} if only, with given density of the liquid, the radius of the field-producing sphere is chosen large enough.

This argument, however, is not convincing; the concept of an incompressible liquid is not compatible with relativity theory as elastic waves would have to travel with infinite velocity. It would be necessary, therefore, to introduce a compressible liquid whose equation of state excludes the possibility of sound signals with a speed in excess of the velocity of light. But the treatment of any such problem would be quite involved; besides, the choice of such an equation of state would be arbitrary within wide limits, and one could not be sure that thereby no assumptions have been made which contain physical impossibilities.

One is thus led to ask whether matter cannot be introduced in such a way that questionable assumptions are excluded from the very beginning. In fact this can be done by choosing, as the field-producing mass, a great number of

small gravitating particles which move freely under the influence of the field produced by all of them together. This is a system resembling a spherical star cluster. Hereby we may proceed as if the field, in which the particles are moving, were produced by a continuous mass distribution of spherical symmetry, corresponding to the whole of the particles.

We can further simplify our considerations by the special assumption that all particles move along circular paths around the center of symmetry of the cluster. Even in this case it is still possible to choose arbitrarily the radial distribution of mass density. The result of the following consideration will be that it is impossible to make g_{44} zero anywhere, and that the total gravitating mass which may be produced by distributing particles within a given radius, always remains below a certain bound.

1. On the paths of the particles and their spacial distribution

By a suitable choice of the radial coördinate, it is possible to obtain the gravitational field of the cluster of spherical symmetry in the form

$$ds^2 = -a(dx_1^2 + dx_2^2 + dx_3^2) + b\,dt^2, \tag{2}$$

whereby a and b are functions of $r = (x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}}$. First we shall investigate the circular motion of one particle around the center of symmetry. Suppose, for instance, this motion takes place within the plane $x_3 = 0$. Through the introduction of polar coördinates

$$x_3 = r \cos \vartheta,$$

$$x_1 = r \sin \vartheta \cos \varphi,$$

$$x_2 = r \sin \vartheta \sin \varphi,$$

(2) assumes the form

$$ds^2 = -a[dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta\, d\varphi^2)] + b\,dt^2. \tag{2a}$$

The field is characterized by

$$g_{11} = -a, \qquad g_{33} = -ar^2 \sin^2 \vartheta,$$

$$g_{22} = -ar^2, \qquad g_{44} = b,$$

where all the rest of the $g_{\mu\nu}$ vanish. The particle under consideration satisfies the equation

$$\frac{d^2 x_\nu}{ds^2} + \Gamma^\nu_{\alpha\beta} \frac{dx_\alpha}{ds} \frac{dx_\beta}{ds} = 0. \tag{3}$$

In addition its motion is determined by the conditions

$$\frac{dx_1}{ds} = \frac{dr}{ds} = 0, \qquad \frac{d^2 x_3}{ds^2} = \frac{d^2 \varphi}{ds^2} = 0,$$

$$x_2 = \vartheta = \frac{\pi}{2}, \qquad \frac{d^2 x_4}{ds^2} = \frac{d^2 t}{ds^2} = 0.$$

It turns out that (3) is satisfied when

$$\Gamma^1_{33}\frac{dx_3}{ds}\frac{dx_3}{ds}+\Gamma^1_{44}\frac{dx_4}{ds}\frac{dx_4}{ds}=0,$$

or when

$$-(ar^2)'\left(\frac{d\varphi}{dt}\right)^2+b=0. \tag{4}$$

Because of (2a), we have

$$\left(\frac{ds}{dt}\right)^2=-ar^2\left(\frac{d\varphi}{dt}\right)^2+b. \tag{5}$$

Thus, $d\varphi/dt$ and ds/dt are determined when the field is given.

Because ds^2 has to be positive for the world line of a particle in motion we have

$$\left(\frac{ds}{dt}\right)^2=b-ar^2\left(\frac{d\varphi}{dt}\right)^2=b-ar^2\frac{b'}{(ar^2)'}>0,$$

or

$$1-\frac{\dfrac{b'}{b}}{\dfrac{(ar^2)'}{ar^2}}>0. \tag{6}$$

By applying this condition to Schwarzschild's field (1) we obtain

$$r>\frac{\mu}{2}(2+\sqrt{3}). \tag{6a}$$

It follows that in the case of a Schwarzschild field a particle is bound to follow a path with a radius greater than $(2+\sqrt{3})$ times the radius of the Schwarzschild singularity. This fact has the greatest significance for the following investigation: In the outermost layer of our particle cluster (and beyond it) the gravitational field is given by (1). It follows that the total gravitating mass of the cluster determines a lower limit for the radius of the cluster; this radius is (in coördinate measure) more than $(2+\sqrt{3})$ times greater than the radius of the Schwarzschild singularity as defined by the field in the empty space outside the cluster.

The normal to the plane in which the particle considered moves has the direction of x_3. If it is assumed that the normals to an infinite number of such planes are distributed at random and also that the phase angles of the paths are subject to a random distribution, then we obtain a cluster of particles of spherical symmetry whose paths have the radius r. The most general cluster to be considered by us consists of an infinite number of clusters of this special type which belong to all values of r. (More accurately speaking, the whole cluster consists, of course, of a finite number of particles so that a field is created which only approximates spherical symmetry.)

In order to formulate the conditions of dynamical equilibrium of the cluster under the influence of its own gravitational field, we first have to compute the energy tensor belonging to such a cluster. For this purpose we assume, for the sake of simplicity that all particles have the same mass m.

2. The Matter-Energy Tensor of the Cluster

We consider the motion of particles within a volume element on the x_3-axis. The velocity vectors all have the same amount, they are perpendicular on the x_3-direction, and they are evenly distributed with respect to the directions within the x_1, x_2 - plane. We know further that the matter-energy tensor depends also on the particle density and on the gravitational potentials, but not on the derivatives of the latter. It is, therefore, possible to determine this tensor by a straightforward calculation.

First we consider particles, with the mass m and the particle density n_0 per unit volume, at rest with respect to a coördinate system of the theory of restricted relativity. In such a case of the energy tensor only the (44)-component exists,

$$T^{44} = mn_0 \frac{dx_4}{ds}\frac{dx_4}{ds}.$$

With respect to coördinate systems in relative motion in the x_1-direction we have the components

$$T^{11} = mn_0 \frac{dx_1}{ds}\frac{dx_1}{ds}, \qquad T^{44} = mn_0 \frac{dx_4}{ds}\frac{dx_4}{ds},$$

$$T^{14} = mn_0 \frac{dx_1}{ds}\frac{dx_4}{ds}.$$

The particle density n with respect to such a system is determined by the equations:

$$n_0V_0 = nV, \qquad V_0\,ds = V\,dt,$$

where V_0 and V denote the rest volume and the coördinate volume respectively. Therefore we have

$$n_0 = n\frac{ds}{dx_4}.$$

We now consider the case when the velocity vector of the particle makes an angle α with respect to the x_1-axis, and is perpendicular to the x_3-axis. By using the relations derived above and by introducing $dl^2 = dx_1^2 + dx_2^2$, we obtain

$$T^{11} = mn\frac{ds}{dx_4}\left(\frac{dl}{ds}\right)^2\cos^2\alpha, \qquad T^{12} = mn\frac{ds}{dx_4}\left(\frac{dl}{ds}\right)^2\cos\alpha\sin\alpha,$$

$$T^{22} = mn\frac{ds}{dx_4}\left(\frac{dl}{ds}\right)^2\sin^2\alpha, \qquad T^{14} = mn\frac{ds}{dx_4}\frac{dl}{ds}\frac{dx_4}{ds}\cos\alpha,$$

$$T^{44} = mn\frac{ds}{dx_4}\left(\frac{dx_4}{ds}\right)^2, \qquad T^{24} = mn\frac{ds}{dx_4}\frac{dl}{ds}\frac{dx_4}{ds}\sin\alpha,$$

all the other components of the energy tensor being zero. In the case that the velocity vectors are evenly distributed over all values of α the result is

$$T^{11} = T^{22} = \tfrac{1}{2}mn \frac{ds}{dx_4}\left(\frac{dl}{ds}\right)^2 = T_{11} = T_{22},$$

$$T^{44} = mn \frac{dx_4}{ds} = T_{44}.$$

We now proceed to the case that the components of the metric tensor are $g_{11} = g_{22} = g_{33} = -a$ and $g_{44} = b$. The components of the energy tensor are obtained by applying the transformation law for tensors and by transforming the coördinates according to

$$dx_a = a^{\frac{1}{2}} d\bar{x}_a$$

$$dx_4 = b^{\frac{1}{2}} d\bar{x}_4 .$$

We obtain

$$\bar{T}_{11} = \left(\frac{dx_a}{d\bar{x}_a}\right)^2 T_{11} = aT_{11},$$

$$\bar{T}_{44} = \left(\frac{dx_4}{d\bar{x}_4}\right)^2 T_{44} = bT_{44}.$$

dl and dx_4, contained in T_{11} and T_{44}, are to be replaced dl by $a^{\frac{1}{2}} dl$ and dx_4 by $b^{\frac{1}{2}} d\bar{x}_4$. Further we have to introduce the particle density with respect to the new coördinates, $\bar{n}$, according to

$$n\, dx_1 dx_2 dx_3 = \bar{n}\, d\bar{x}_1 d\bar{x}_2 d\bar{x}_3$$

or

$$n = \bar{n} a^{-\frac{3}{2}}.$$

After having made all these transformations and substitutions, and omitting the bars denoting the new coördinate system, we obtain

$$\text{(7)} \quad \begin{cases} T_{11} = T_{22} = \tfrac{1}{2} mna^{\frac{1}{2}} b^{-\frac{1}{2}} \dfrac{ds}{dx_4}\left(\dfrac{dl}{ds}\right)^2, \\[2ex] T_{44} = mna^{-\frac{3}{2}} b^{\frac{3}{2}} \dfrac{dx_4}{ds}. \end{cases}$$

In these equations ds/dx_4 and dl/ds have to be replaced by the expressions given by (4) and (5) which were derived from the equations of the geodesic lines. Further we write dt instead of dx_4 and $r\,d\varphi$ instead of dl. The final result is

$$\text{(7a)} \quad \begin{cases} T_{11} = T_{22} = \tfrac{1}{2} mna^{-\frac{1}{2}} \dfrac{\beta'}{\alpha'}\left(\dfrac{\alpha'}{\alpha' - \beta'}\right)^{\frac{1}{2}}, \\[2ex] \dfrac{a}{b} T_{44} = mna^{-\frac{1}{2}}\left(\dfrac{\alpha'}{\alpha' - \beta'}\right)^{\frac{1}{2}}, \end{cases}$$

where α and β denote the expressions

$$\begin{cases} \alpha = \ln (r^2 a), \\ \beta = \ln b. \end{cases} \tag{7b}$$

3. The Differential Equations of the Gravitational Field

The differential equation of a gravitational field which is due to a matter-energy tensor are

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \kappa T_{\mu\nu} = 0. \tag{8}$$

These equations have to be specialized for a static field of the type (2). By a straight forward calculation the following equations are obtained for a point on the x_3-axis:

$$-G_{33} = \frac{a'}{ra} + \frac{b'}{rb} + \frac{1}{4}\left(\frac{a'}{a}\right)^2 + \frac{1}{2}\frac{a'}{a}\frac{b'}{b} = 0, \tag{9}$$

$$G_{11} = -\frac{1}{2}\left(\frac{a'}{a}\right)' - \frac{1}{2}\left(\frac{b'}{b}\right)' - \frac{1}{2}\frac{a'}{ra} - \frac{1}{2}\frac{b'}{rb} - \frac{1}{4}\left(\frac{b'}{b}\right)^2 + \kappa T_{11} = 0, \tag{10}$$

$$\frac{a}{b} G_{44} = \left(\frac{a'}{a}\right)' + 2\frac{a'}{ra} + \frac{1}{4}\left(\frac{a'}{a}\right)^2 + \kappa T_{44}\frac{a}{b} = 0. \tag{11}$$

For T_{11} and T_{44} we have to substitute the expressions given by (7a), (7b). As m is to be considered a given constant, the only functions of the coördinates in these equations are n, a, and b. It is to be expected in the first place that n, i.e. radial distribution of matter, remains undetermined by the equations. This makes necessary the existence of an identity between the equations (9), (10), (11). In fact such an identity exists. Its form is

$$0 \equiv G'_{33} + \left(\frac{2}{r} + \frac{1}{2}\frac{b'}{b}\right) G_{33} - \left(\frac{2}{r} + \frac{a'}{a}\right) G_{11} + \frac{1}{2}\frac{b'}{b} G_{44}. \tag{12}$$

It may be obtained in the following way: We have constructed $T_{\mu\nu}$ by considering particles which satisfy the equations of motion in the field. Therefore the covariant divergence of this tensor is bound to vanish identically. On the other hand, the divergence of $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ vanishes identically on account of the Bianchi identities. Of these four equations having the form of divergences only the one with the index 3 yields anything which does not already vanish identically with respect to the $G_{\mu\nu}$, and that is (12). From the form of (12) it follows that (10) is the consequence of (9) and (11). The problem is therefore reduced to (9) and (11), and the particle density remains undetermined, as was to be expected.

This result makes possible a further simplification of the problem. If, in (9), the quantities $\alpha = \ln (r^2 a)$ and $\beta = \ln b$ are introduced, we obtain the equation

$$-\frac{2}{r^2} + \frac{1}{2}\alpha'^2 + \alpha'\beta' = 0. \tag{13}$$

By taking into account (13) and (7a), we obtain from (11)

$$\alpha'' + \frac{\alpha'}{r} + \frac{1}{4}\alpha'^2 - \frac{1}{r^2} + \kappa mna^{-\frac{1}{2}}\left(\frac{\alpha'^2}{\frac{3}{2}\alpha'^2 - \frac{2}{r^2}}\right)^{\frac{1}{2}} = 0. \tag{14}$$

This is a differential equation for a alone. When a is already known b is obtained by a simple integration from

$$\beta' = \frac{1}{\alpha'}\left(\frac{2}{r^2} - \frac{1}{2}\alpha'^2\right). \tag{13a}$$

4. Localization of the Particles within a Thin Spherical Shell

Outside the cluster, the gravitational field is represented by Schwarzschild's solution which, with our choice of the coördinate system, is given by (1). Inside the cluster, the field is determined by (14). Thereby, the function n is to be considered as given. However, n is not completely arbitrary, as the total radius of the cluster is restricted by the lower limit given by (6a).

Equation (14) represents a complicated relation between the particle density n and the function a representing the gravitational field. The limiting case, however, in which the gravitating particles are concentrated within an infinitely thin spherical shell, between $r = r_0 - \Delta$ and $r = r_0$, is comparatively simple. Of course, this case could only be realized if the individual particles had the rest-volume zero, which cannot be the case. This idealization, however, still is of interest as a limiting case for the radial distribution of the particles.

We divide the whole space into three zones for separate consideration, part O to be the part outside the shell, $r \geqq r_0$, part I to be the part inside the shell, $r \leqq r_0 - \Delta$, and part S to be the part of the shell $r_0 - \Delta \leqq r \leqq r_0$. In O, the gravitational field is represented by (1), in I, it is represented by (2) with constant values of a and b. It follows that a' (and α') have to change within S the faster the smaller Δ is chosen. However, as a' remains finite in S, a itself changes only infinitely little in S. It is, therefore, permissible in S to neglect α' compared with α''. We therefore replace (14) within S by

$$\alpha'' + \kappa mna^{-\frac{1}{2}}\left(\frac{\alpha'^2}{\frac{3}{2}\alpha'^2 - \frac{2}{r^2}}\right)^{\frac{1}{2}} = 0, \tag{14a}$$

where a and r are to be treated as constants for integration purposes. We introduce the variable

$$z^2 = \tfrac{3}{4} r^2\alpha'^2 - 1$$

and the "constant"

$$C = \kappa ma^{-\frac{1}{2}} \frac{r}{\sqrt{2}}$$

and obtain the equation

$$\left(1 - \frac{1}{1 + z^2}\right) dz = Cn\,dr. \tag{14b}$$

z is hereby determined as a function of r within S if n is given as a function of r. When the integration is carried out between $r_0 - \Delta$ and r_0 we obtain

$$| z - \operatorname{arctg} z |_{r_0}^{r_0-\Delta} = \frac{C}{4\pi r_0^2} N = \frac{\kappa}{8\pi} \sqrt{2}\, a^{-\frac{1}{2}} \frac{mN}{r_0}, \tag{15}$$

where N designates the number of particles in S. It follows from (1) that for $r = r_0$

$$z_{r_0} = \sqrt{2}\, \frac{(1 - 4\sigma + \sigma^2)^{\frac{1}{2}}}{1 + \sigma}, \qquad \sigma = \frac{\mu}{2r_0}, \tag{15a}$$

and from (2) that, because of a and b being constant in I, in I

$$z_{r_0-\Delta} = \sqrt{2}. \tag{15b}$$

It follows from (6a) that

$$\sigma < \frac{1}{2 + \sqrt{3}} = 2 - \sqrt{3}.$$

It turns out that this is just the condition for the numerator of the expression for z_{r_0} to be real. (15), for each possible r_0, gives the relationship between the sum of the masses of the particles, mN, and the total gravitating mass μ of the cluster. For large values of r_0, with a fixed value of μ, one obtains in the limit

$$\mu = \frac{\kappa}{8\pi}\, mN. \tag{16}$$

The factor $\kappa/8\pi$ is due to the fact that m is measured in grams, μ, however, in gravitational units. (16) therefore simply states that in this limiting case the gravitating mass of the cluster is equal to the sum of the particle masses.

The most illuminating way to express this result is the following:

Outside the shell ($r \geqq r_0$), the gravitational field is given by

$$ds^2 = -\left(1 + \frac{\mu}{2r}\right)^4 (dx_1^2 + dx_2^2 + dx_3^2) + \frac{1 - \frac{\mu}{2r}}{1 + \frac{\mu}{2r}}\, dt^2.$$

Inside the shell it is given by the same expression, with the difference, however, that r is to be replaced by the constant r_0, whereby the inequality

$$r_0 > \frac{\mu}{2}(2 + \sqrt{3})$$

must be satisfied. The number N of particles of the mass m which together form the shell is given by the following consideration: As an abbreviation we introduce

$$\Sigma = \frac{\kappa}{8\pi}\frac{mN}{2r_0} = \frac{M}{2r_0}, \qquad \sigma = \frac{\mu}{2r_0}.$$

Then we have

$$\Sigma = \phi(\sigma) = \frac{[(\sqrt{2} - \operatorname{arctg}\sqrt{2}) - (z_{r_0} - \operatorname{arctg} z_{r_0})](1+\sigma)^2}{\sqrt{8}},$$

where

$$z_{r_0} = \sqrt{2}\,\frac{(1 - 4\sigma + \sigma^2)^{\frac{1}{2}}}{1+\sigma}.$$

σ can assume values between 0 and $2 - \sqrt{3}(\sim .27)$. The quantity

$$\frac{\Sigma - \sigma}{\sigma}$$

is only very little different from zero in this whole region. A few typical values are given in the following table:

σ	$\frac{\Sigma - \sigma}{\sigma}$
.05	.042
.14	.06
.2	.055
.23	.013
.27	−0.022

This leads to a very interesting consequence: First it is clear that $(\Sigma - \sigma)/\sigma$ may be replaced by $(\Sigma - \sigma)/\Sigma$ with good approximation and this by $(M - \mu)/M$. This latter quantity is the relative decrease of energy of the cluster when it contracts from an infinite radius to the radius r_0. The table shows that this contraction energy has a maximum near $\sigma = 0.15$, and for greater values of σ, i.e. smaller values of r_0, it decreases again. The physical cause of this effect is that, with decreasing r_0, the potential energy of the cluster decreases, but the kinetic energy increases. For sufficiently small values of r_0 the latter effect surpasses the former.

It is therefore clear that the decrease of the radius with decreasing energy would come to an end for a value of about $\sigma = 0.15$, i.e. a radius of about $6.7(\mu/2r_0)$, while the lower limit of the radius as given by the velocity of light is $(2 + \sqrt{3})\ (\mu/2r_0)$. The value of r corresponding to the minimum energy means an upper limit for the particle velocity in the direction of the tangent of about 0.65 times the light velocity.

5. Qualitative Discussion of the Case of Arbitrary Radial Mass Distribution

We consider the case of a given mass μ and a shell radius r_0 satisfying the inequality (6a). When a number N of particles is brought into this shell zone,

as determined by (15), then the exterior gravitational field is just completely screened off from the interior I so that there the field will be Euclidean. This means that the line element in I is characterized by constant values of a and b, where b cannot reach its lower limit $1/\sqrt{3}$.

If, however, the number of particles in S is chosen smaller than according to (15) then the field will not be screened off entirely (μ is hereby regarded as being kept fixed). We can then satisfy the theory formally by replacing the Euclidean line element in I by a Schwarzschild line element of the form

$$a = A\left(1 + \frac{\mu_1}{2r}\right)^4, \qquad b = B\left(\frac{1 - \frac{\mu_1}{2r}}{1 + \frac{\mu_1}{2r}}\right)^2,$$

where A, B, and μ_1 are constants. μ_1 will be smaller than μ which characterizes the field outside the shell. This interior field has a singularity of the Schwarzschild type ($b = 0$) at $r = \mu_1/2$.

This singularity, however, can be removed by introducing a second shell S_1 inside S, which has to be constructed so that the gravitational field in its interior will be Euclidean. The whole cluster will then consist of two shells S and S_1 and will have no Schwarzschild singularity.

Again this system can be modified by reducing the number of particles in S_1 so that it will not screen off *its* exterior field (between S and S_1) entirely; then a third shell S_2, of still smaller radius, may be constructed so that *its* exterior field is just screened off entirely from its interior.

This method can be reiterated up to the center of the cluster. Thus one obtains clusters with the most varied radial mass distributions. There will be also various steady distributions. It is impossible, however, that b should vanish anywhere. The radius of the cluster will always be greater than the limiting radius $\frac{1}{2}\mu(2 + \sqrt{3})$, and it will not be possible to concentrate the matter of the cluster arbitrarily densely near the center of the cluster.

6. The Case of Continuous Particle Density

The consideration given in part 5. leads toward the solution for continuous distributions of the particle density. We divide the interval $0 \leqq r \leqq r_0$ into an infinite number of equal parts dr. We imagine that there is constructed in the center of each partition dr a shell of a two dimensional character of the type discussed in part 4. The shells may be chosen so that they are equivalent to a continuous distribution of mass. Between any two subsequent shells we shall have a gravitational field of the Schwarzschild type

$$ds^2 = -A\left(1 + \frac{\tau}{2r}\right)^4 (dx_1^2 + dx_2^2 + dx_3^2) + B\left(\frac{1 - \frac{\tau}{2r}}{1 + \frac{\tau}{2r}}\right)^2 dt^2, \tag{17}$$

where A, B, and τ are constants which differ only infinitesimally for two neighboring regions. Then the sum total of all these partial solutions constitutes the

gravitational field inside the cluster. Our task is to determine A, B, and τ as functions of r.

We consider two neighboring Schwarzschild solutions which belong to the radius intervals $r - \frac{1}{2}dr$ to $r + \frac{1}{2}dr$ and $r + \frac{1}{2}dr$ to $r + \frac{3}{2}dr$. In the first region the values of A, B, and τ belong to the value r of the radius, in the second to the value $r + dr$. If we use the quantities introduced by (2) then the two local solutions are given by

$$a(r; A, \tau), \qquad a(r; A + dA, \tau + d\tau),$$

and

$$b(r; B, \tau), \qquad b(r; B + dB, \tau + d\tau),$$

where a, b are functions of r in accordance with (17). These two solutions are to assume the same values for a and b in the point $r + \frac{1}{2}dr$ because these quantities must not change when we pass through a shell occupied by particles. It follows, up to quantities of the first order

$$\frac{\partial a}{\partial A} dA + \frac{\partial a}{\partial \tau} d\tau = 0,$$

$$\frac{\partial b}{\partial B} dB + \frac{\partial b}{\partial \tau} d\tau = 0,$$

or, in accordance with (17)

$$\text{(18)} \qquad \begin{cases} \dfrac{dA}{A} + \dfrac{4}{r} \dfrac{r\,d\sigma + \sigma\,dr}{1 + \sigma} = 0, \\[2ex] \dfrac{dB}{B} - \dfrac{4}{r} \dfrac{r\,d\sigma + \sigma\,dr}{(1 + \sigma)(1 - \sigma)} = 0, \end{cases}$$

where σ is written for $\tau/2r$.

These equations determine A, B as functions of r when τ or σ is given as function of r. It turns out that α, β, computed from the solutions A, B of (18), are the solutions of (13), represented with the help of the "parameter" function σ. τ is arbitrary within certain limits because it is closely connected with the mass distribution. On the other hand, A, B, and τ have to satisfy the condition that (17) makes possible circular particle paths for all values of r, i.e. a and b have to satisfy the inequality (6). In connection with (17) we obtain the inequality

$$\text{(19)} \qquad 1 - \frac{\dfrac{B'}{B} - 4\,\dfrac{\sigma'}{(1 + \sigma)(1 - \sigma)}}{\dfrac{A'}{A} + \dfrac{2}{r} + 4\,\dfrac{\sigma'}{1 + \sigma}} > 0.$$

(18) and (19) together completely determine the problem within the cluster; σ is arbitrary save for the only restriction that, together with the values of A and B, calculated from (18), it has to satisfy (19).

For $r \geqq r_0$ we have, of course, $A = B = 1$, with $\tau = \text{const.} = \mu$.

By using (18) we may write (19) thus:

$$1 - \frac{4 \dfrac{\sigma}{(1+\sigma)(1-\sigma)}}{2 - 4\dfrac{\sigma}{1+\sigma}} > 0$$

or, with some transformations:

$$\text{(19a)} \qquad \frac{(\sigma - 2 + \sqrt{3})(\sigma - 2 - \sqrt{3})}{(1-\sigma)^2} > 0.$$

This inequality has to hold within as well as outside the cluster. For infinite values of r, σ vanishes. Further σ has to be positive, as negative masses are excluded. Because of the denominator, σ can nowhere be greater than 1. Therefore the numerator of the left hand side has to be positive. As the second factor of the numerator is always negative the first factor has to be negative, too. We therefore obtain

$$\text{(19b)} \qquad \sigma < 2 - \sqrt{3}.$$

This is a generalization of (6a) as (6a) was only proven to hold for the outside boundary of the cluster.

τ represents the mass enclosed by the spherical surface of the radius r. In order that negative masses should be ruled out it is necessary that everywhere

$$\text{(20)} \qquad \frac{d\tau}{dr} \geqq 0.$$

It is further necessary that τ vanishes for $r = 0$. Save for this condition τ may be chosen arbitrarily if only σ satisfies (19b). When τ and therefore σ is given then the problem of determining the gravitational field of the form (17) is reduced to the carrying out of two integrations, according to (18).

The equations (18) give us the integration of (13) with arbitrary mass density distribution, where the latter is expressed by τ or σ. (14) gives the correspond-ign particle density n. We shall express n in terms of σ. We have

$$\text{(21)} \qquad 0 = \frac{2}{r}\left(\frac{1-\sigma}{1+\sigma}\right)' - \frac{4}{r^2}\frac{\sigma}{(1+\sigma)^2} + \frac{1}{2}\kappa m n a^{-\frac{1}{2}} \frac{1-\sigma}{\sqrt{1 - 4\sigma + \sigma^2}}$$

together with the relations

$$\text{(22)} \qquad a = A(1+\sigma)^4, \qquad \frac{A'}{A} = -\frac{4}{r}\frac{r\sigma' + \sigma}{1+\sigma}.$$

Therefore, when σ is given as a function of r we obtain n by carrying out one integration only.

σ is positive and stays below the limit $2 - \sqrt{3}$. The square root of the denominator of the third term in (21) therefore is always positive. We further

have $\tau/2r$ where τ is the gravitating mass contained in a sphere of the radius r. τ therefore increases monotonically with increasing r. If the mass density is to be finite in the region around $r = 0$ then τ has to decrease in that region at least as fast as r^3 and σ at least as fast as r^2. Under these conditions the two first terms in (21) will be finite everywhere, and also A'/A, A, and a. (21) therefore gives us a finite value for n. It is further possible to prove from the properties of τ that the sum of the two first terms in (21) is negative everywhere.

From all these considerations it can be followed that a and b are finite and not zero in the whole space.

By combining (2), (4), (17), and (18) one can show that the ratio V between the particle velocity and between the light velocity pointing into the same direction, is given by

$$V^2 = \frac{\beta'}{\alpha'} = \frac{2\sigma}{(1-\sigma)^2}. \tag{23}$$

When σ stays below a given limit V will stay below a certain limit, too.

7. A Special Case of Continuous Mass Distribution

It is of some interest to investigate the case where σ inside the cluster is a constant σ_0. Strictly speaking this case falls outside of our conditions as σ ought to decrease toward the point $r = 0$ at least as fast as r^2 in order that the density in the neighborhood of the center should stay finite. We can satisfy this condition by choosing σ for instance

$$\sigma = \sigma_0(1 - e^{-cr^2}) \tag{24}$$

where c is to be an arbitrary constant. We then consider from the start the limiting case of $c = \infty$. This special case is discussed here in order to supplement the discussions of part 4. There the whole mass was distributed as far outside (within the total radius r_0) as possible, while here we have a strong concentration of mass toward the center of the cluster.

As τ is the gravitating mass enclosed by a spherical surface of the radius r, $d\tau/(4\pi r^2 dr)$ is the mean density of the gravitating mass in the point r. As $\tau = 2\sigma_0 r$ we obtain for this mean density $\sigma_0/2\pi r^2$, i.e. a radial decrease of the density like $1/r^2$ up to the cluster boundary $r = r_0$.

From (18), in accordance with (24) (in the limiting case of vanishing expo nential term), we obtain

$$\left\{\begin{aligned} \frac{dA}{A} &= -\frac{4\sigma_0}{1+\sigma_0}\frac{dr}{r}, \\ \frac{dB}{B} &= \frac{4\sigma_0}{1-\sigma_0^2}\frac{dr}{r}, \end{aligned}\right. \tag{18a}$$

and since for $r = r_0$, A and B have to assume the value 1

$$\text{(18b)} \qquad \begin{cases} A = \left(\dfrac{r}{r_0}\right)^{-4\sigma_0/(1+\sigma_0)}, \\ B = \left(\dfrac{r}{r_0}\right)^{4\sigma_0/(1-\sigma_0^2)}. \end{cases}$$

For $r = 0$ we obtain $a = \infty$ and $b = 0$. This type of singularity, however, is not to be taken seriously because it would be avoided if we had taken into consideration the exponential term in (24). It is to be noted that through a suitable choice of the mass distribution this singularity can be approximated, but not reached.

We make use of (21) in order to determine the relation existing between the sum of the rest masses of the particles M

$$M = \frac{\kappa}{8\pi} m \int_0^{r_0} n \cdot 4\pi r^2 \, dr,$$

and the total gravitating mass of the cluster μ. It can be shown that the first term of (21) gives only a vanishing contribution for infinitely great values of c. This follows from the fact that $\left(\dfrac{1-\sigma}{1+\sigma}\right)'$ vanishes everywhere where the influence of the exponential term of (24) has become unnoticeable. We compute the contribution of the second term in (21) by omitting the exponential term from the start and obtain, after a short calculation, as the final result, with $\mu = 2r_0\sigma_0$

$$\text{(25)} \qquad M = \mu(1 - 4\sigma_0 + \sigma_0^2)^{\frac{1}{2}} \frac{1 + \sigma_0}{(1 - \sigma_0)^2}.$$

This equation when compared with the relation

$$\text{(26)} \qquad \mu = 2\sigma_0 r_0$$

allows an easy discussion of the essential properties of clusters of this type.

First it is easy to see that we have extremely simple relations when we change M but keep fixed σ_0 ($0 < \sigma_0 < 2 - \sqrt{3}$) and thereby the tangential velocity of the particles as measured in light velocity units. When M is multiplied by z the gravitating mass will be $z\mu$ and the diameter of the cluster will be $z \cdot 2r$. The mean density will be multiplied by $1/z^2$.

In order to obtain a survey of all possibilities it is therefore sufficient to keep fixed the number of constituting particles and thereby M and to vary σ_0 together with the diameter $2r_0$ and the gravitating mass μ. We obtain for $M = 1$

$$\mu = \frac{(1 - \sigma_0)^2}{1 + \sigma_0} (1 - 4\sigma_0 + \sigma_0^2)^{-\frac{1}{2}}.$$

The following table gives μ and $2r_0$ for $M = 1$ as functions of σ_0 (approximately):

σ_0	μ	$2r_0$
0.	1.	∞
.05	.988	19.76
.1	.948	9.48
.15	.97	6.56
.2	1.13	5.65
.23	1.32	5.63
.25	1.82	7.40
.26	2.63	10.1
.268	∞	∞

When the cluster is contracted from an infinite diameter its mass decreases at the most about 5%. This minimal mass will be reached when the diameter $2r_0$ is about 9. The diameter can be further reduced down to about 5.6, but only by adding enormous amounts of energy. It is not possible to compress the cluster any more while preserving the chosen mass distribution. A further addition of energy enlarges the diameter again. In this way the energy content, i.e. the gravitating mass of the cluster, can be increased arbitrarily without destroying the cluster. To each possible diameter there belong two clusters (when the number of particles is given) which differ with respect to the particle velocity.

Of course, these paradoxical results are not represented by anything in physical nature. Only that branch belonging to smaller σ_0 values contains the cases bearing some resemblance to real stars, and this branch only for diameter values between ∞ and $9M$.

The case of the cluster of the shell type, discussed earlier in this paper, behaves quite similarly to this one, despite the different mass distribution. The shell type cluster, however, does not contain a case with infinite μ, given a finite M.

The essential result of this investigation is a clear understanding as to why the "Schwarzschild singularities" do not exist in physical reality. Although the theory given here treats only clusters whose particles move along circular paths it does not seem to be subject to reasonable doubt that more general cases will have analogous results. The "Schwarzschild singularity" does not appear for the reason that matter cannot be concentrated arbitrarily. And this is due to the fact that otherwise the constituting particles would reach the velocity of light.

This investigation arose out of discussions the author conducted with Professor H. P. Robertson and with Drs. V. Bargmann and P. Bergmann on the mathematical and physical significance of the Schwarzschild singularity. The problem quite naturally leads to the question, answered by this paper in the negative, as to whether physical models are capable of exhibiting such a singularity.

THE INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 40, No. 4, October, 1939

TENSOR EQUATIONS EQUIVALENT TO THE DIRAC EQUATIONS

By A. H. Taub

(Received November 1, 1938; revised May 5, 1939)

1. The Spinor Formalism

In this paper the two component spinor formalism will be used to study the effect of geometric spin transformations corresponding to both proper and improper Lorentz transformations on the tensor equations which Whittaker[1] and Ruse[2] have shown to be equivalent to the Dirac equations. The spinor form of the Maxwell equations given by Laporte and Uhlenbeck[3] is obtained with the aid of the formalism. These equations are compared with the equations proposed by Dirac[4] and used by Kemmer[5] for material particles with spin one. The tensor equations equivalent to these are obtained and are shown to differ from the equations proposed by Proca.[6] Throughout this paper we shall use the notation developed by O. Veblen in a seminar conducted by him and W. Givens.[7]

The Dirac equations when written in two component form are

$$
\begin{aligned}
&g^{\sigma A}{}_{B}\left(\frac{h}{i}\frac{\partial}{\partial x^{\sigma}} - \frac{e}{c}\varphi_{\sigma}\right)\psi^{B} = -imc\bar{\varphi}^{A},\\
(1.1)\qquad &\\
&g^{\sigma A}{}_{B}\left(\frac{h}{i}\frac{\partial}{\partial x^{\sigma}} + \frac{e}{c}\varphi_{\sigma}\right)\varphi^{B} = -imc\bar{\psi}^{A},
\end{aligned}
$$

where m and e are the mass and charge of the electron, h is Planck's constant divided by 2π, c is the velocity of light, φ_σ the electromagnetic four vector potential, and ψ^A, φ^A, and $g^{\sigma A}{}_B$ are spinors. The components of the spinor $g^{\sigma A}{}_B$ depend on the coördinate system in two spaces, namely, the spin space and the space of

---

[1] E. T. Whittaker: "On the Relations of the Tensor-Calculus to the Spinor-Calculus," Proc. Roy. Soc. *158A*, pp. 38–46 (1937).

[2] H. S. Ruse: "On the Geometry of Dirac's Equations and their Expression in Tensor Form," Proc. Roy. Soc. of Edin., LVII, part II, pp. 97–127 (1936–37).

[3] O. Laporte and G. E. Uhlenbeck: "Application of Spinor Analysis to the Maxwell and Dirac Equations," Phys. Rev. 37, p. 1380 (1931).

[4] P. A. M. Dirac: "Relativistic Wave Equations," Proc. Roy. Soc. *A 155*, p. 447 (1936).

[5] N. Kemmer: "Quantum Theory of Einstein-Bose Particles and Nuclear Interaction," Proc. Roy. Soc. *A 166*, p. 127 (1937).

[6] A. Proca: "Sur La Theorie Ondulatoire Des Elections Positifs et Negatifs," J. de Physique et de Radium 7, p. 347 (1936).

[7] O. Veblen and J. Von Neumann: "Geometry of Complex Domains," Princeton Mimeographed notes (1935–36). Notes by W. Givens and A. H. Taub. References to these notes will be denoted by "notes."

special relativity S_4 with coördinates $x^\sigma (x^4 = ct)$ and a metric tensor $g_{\sigma\tau}$ which in a preferred coördinate system has the components

$$\|g_{\sigma\tau}\| = \begin{Vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{Vmatrix}. \tag{1.2}$$

The spinor $g^{\sigma\dot{A}}{}_B$ is related to the tensor $g_{\sigma\tau}$ by the matrix equations

$$\tfrac{1}{2}(\bar{g}^\sigma g^\tau + \bar{g}^\tau g^\sigma) = -g^{\sigma\tau}1, \tag{1.3}$$

where $g^{\sigma\tau}$ is determined by $g_{\sigma\tau}$ in the usual manner, 1 is the identity matrix and $g^\sigma = \| g^{\sigma\dot{A}}{}_B \|$ and the bar denotes the complex conjugate. A solution of these equations is given by the matrices

$$\begin{aligned} -g^1 = g_1 &= \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}, & -g^2 = g_2 &= \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}, \\ -g^3 = g_3 &= \begin{Vmatrix} -i & 0 \\ 0 & i \end{Vmatrix} & \text{and} \quad g^4 = g_4 &= \begin{Vmatrix} 0 & -1 \\ 1 & 0 \end{Vmatrix}. \end{aligned} \tag{1.4}$$

If $\psi = \begin{Vmatrix} \psi^1 \\ \psi^2 \end{Vmatrix}$ and $\varphi = \begin{Vmatrix} \varphi^1 \\ \varphi^2 \end{Vmatrix}$ equations (1.1) may be written as

$$\begin{aligned} g^\sigma p_\sigma \psi &= -imc\bar{\varphi}, \\ \bar{g}^\sigma p_\sigma \bar{\varphi} &= -imc\psi, \end{aligned} \tag{1.5}$$

where $p_\sigma = \dfrac{h}{i}\dfrac{\partial}{\partial x^\sigma} - \dfrac{e}{c}\varphi_\sigma$. From this form of the equations it is readily verified by using (1.3) that both φ and ψ satisfy second order wave equations.

The tensor $g_{\sigma\tau}$ and the tensor $g^{\sigma\tau}$ may be used to raise and lower tensor indices. The spinor indices are raised and lowered by antisymmetric spinors ϵ_{AB} and ϵ^{AB} of weight -1 and $+1$ respectively. In all coördinate systems they have the components

$$\|\epsilon_{AB}\| = \|\epsilon^{AB}\| = \begin{Vmatrix} 0 & 1 \\ -1 & 0 \end{Vmatrix}. \tag{1.6}$$

The rules for raising and lowering spin indices are given by the equations

$$\psi_A = \epsilon_{AB}\psi^B \quad \text{or} \quad \psi_1 = \psi^2, \quad \psi_2 = -\psi^1, \tag{1.7}$$

and

$$\psi^A = \epsilon^{BA}\psi_B. \tag{1.8}$$

Since the spinors ϵ_{AB} and ϵ^{BA} are antisymmetric, the rules for raising are different from those for lowering an index. The rules given by equations (1.6) and (1.7) are consistent. They can be used on spinors with a number of indices. Dotted

indices are manipulated by spinors $\epsilon_{\dot{A}\dot{B}} = \bar{\epsilon}_{AB}$ and $\epsilon^{\dot{A}\dot{B}} = \bar{\epsilon}^{AB}$. As a consequence of our rules we have

$$\psi^A \varphi_A = \psi^A \epsilon_{AB} \varphi^B = -\varphi^A \psi_A, \tag{1.9}$$

hence

$$\psi^A \psi_A = 0. \tag{1.10}$$

If we apply the process of raising and lowering indices to the spinor $g_\sigma{}^{\dot{A}}{}_B$, we obtain two new spinors, $g_\sigma{}^{\dot{A}B}$ and $g_{\sigma \dot{A}B}$. In the coördinate systems in which (1.4) hold we have:

$$\begin{aligned} \| g_{1\dot{A}B} \| &= \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}, & \| g_{2\dot{A}B} \| &= \begin{Vmatrix} 0 & -1 \\ -1 & 0 \end{Vmatrix}, \\ \| g_{3\dot{A}B} \| &= \begin{Vmatrix} 0 & -i \\ i & 0 \end{Vmatrix} \quad \text{and} & \| g_{4\dot{A}B} \| &= \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}, \end{aligned} \tag{1.11}$$

$$\begin{aligned} \| g_1{}^{\dot{A}B} \| &= \begin{Vmatrix} -1 & 0 \\ 0 & 1 \end{Vmatrix}, & \| g_2{}^{\dot{A}B} \| &= \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}, \\ \| g_3{}^{\dot{A}B} \| &= \begin{Vmatrix} 0 & -i \\ i & 0 \end{Vmatrix} \quad \text{and} & \| g_4{}^{\dot{A}B} \| &= \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}. \end{aligned} \tag{1.12}$$

From these equations it is readily verified that

$$\bar{g}_{\sigma \dot{A}B} = g_{\sigma \dot{B}A} \quad \text{and} \quad \bar{g}_\sigma{}^{\dot{A}B} = g_\sigma{}^{\dot{B}A}, \tag{1.13}$$

that is, each of the matrices $\| g_\sigma{}^{\dot{A}B} \|$ and $\| g_{\sigma \dot{A}B} \|$ is Hermitian. Equations (1.13) are spinor equations and hence hold in all coördinate systems, since we assume that the spinor $g^{\sigma \dot{A}B}$ has weight and anti-weight equal to $\frac{1}{2}$ and all contravariant simple spinors are of weight $+\frac{1}{2}$ and covariant ones of weight $-\frac{1}{2}$. (See Chapter II §3 of notes.)

Equations (1.3) may be written as

$$\tfrac{1}{2}(g^{\sigma \dot{A}B} g_{\tau \dot{A}C} + g_\tau{}^{\dot{A}B} g^\sigma{}_{\dot{A}C}) = \delta^\sigma_\tau \delta^B_C. \tag{1.14}$$

By setting $B = C$ and summing, we obtain

$$g^{\sigma \dot{A}B} g_{\tau \dot{A}B} = 2\delta^\sigma_\tau. \tag{1.15}$$

Another identity satisfied by the components of the spinor $g^{\sigma \dot{A}B}$ is

$$g^{\sigma \dot{A}B} g_{\sigma \dot{C}D} = 2\delta^{\dot{A}}_{\dot{C}} \delta^B_C, \tag{1.16}$$

as may be verified by using the components of the spinor in the particular coördinate systems. Equations (1.15) and (1.16) are the analytical consequences of the fact that there is a (1-1) correspondence between Hermitian second order matrices and points of S_4 and may be derived from that correspondence.

The (1-1) correspondence between self dual antisymmetric tensors in S_4 and

symmetric spinors may be studied by using the spinor $S_{\sigma\tau}{}^{A}{}_{C}$ defined by the equations

$$(1.17) \qquad S_{\sigma\tau}{}^{A}{}_{C} = \tfrac{1}{2}(g_{\sigma}{}^{\dot{B}A} g_{\tau\dot{B}C} - g_{\tau}{}^{\dot{B}A} g_{\sigma\dot{B}C}) = \tfrac{1}{2}(\bar{g}_{\tau}{}^{\dot{A}}{}_{B} g_{\sigma}{}^{\dot{B}}{}_{C} - \bar{g}_{\sigma}{}^{\dot{A}}{}_{B} g_{\tau}{}^{\dot{B}}{}_{C}).$$

From equation (1.17) it follows that

$$(1.18) \qquad S_{\sigma\tau}{}^{A}{}_{A} = 0 \quad \text{or} \quad S_{\sigma\tau AB} = S_{\sigma\tau BA}.$$

From equation (1.14) and (1.17) we find

$$(1.19) \qquad g_{\sigma}{}^{\dot{B}A} g_{\tau\dot{B}C} = S_{\sigma\tau}{}^{A}{}_{C} + g_{\sigma\tau}\delta^{A}{}_{C}.$$

By direct application of equation (1.16) we obtain the following useful equations:

$$(1.20) \qquad S^{\sigma\tau AB} S_{\sigma\tau CD} = 4(\delta_{C}^{A}\delta_{D}^{B} + \delta_{D}^{A}\delta_{C}^{B}),$$

$$(1.21) \qquad \bar{S}^{\sigma\tau AB} S_{\sigma\tau CD} = 0,$$

$$(1.22) \qquad g^{\sigma\dot{C}D} S_{\sigma\tau AB} = -\delta_{A}^{D} g_{\tau}{}^{\dot{C}}{}_{B} - \delta_{B}^{D} g_{\tau}{}^{\dot{C}}{}_{A},$$

and

$$(1.23) \qquad S^{\sigma\tau}{}_{AB} \bar{S}_{\tau\rho}{}^{EF} = \tfrac{1}{2}[g^{\sigma\dot{E}}{}_{A} g_{\rho}{}^{\dot{F}}{}_{B} + g^{\sigma\dot{F}}{}_{A} g_{\rho}{}^{\dot{E}}{}_{B} + g^{\sigma\dot{E}}{}_{B} g_{\rho}{}^{\dot{F}}{}_{A} + g^{\sigma\dot{F}}{}_{B} g_{\rho}{}^{\dot{E}}{}_{A}].$$

From the second of equations (1.17) and (1.4) we find that in the special coördinate systems

$$(1.24) \qquad \begin{aligned} \| S_{14}{}^{A}{}_{B} \| &= \begin{Vmatrix} -1 & 0 \\ 0 & 1 \end{Vmatrix}, & \| S_{12}{}^{A}{}_{B} \| &= \begin{Vmatrix} 0 & 1 \\ -1 & 0 \end{Vmatrix}, \\ \| S_{24}{}^{A}{}_{B} \| &= \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}, & \| S_{13}{}^{A}{}_{B} \| &= \begin{Vmatrix} 0 & i \\ i & 0 \end{Vmatrix}, \\ \| S_{34}{}^{A}{}_{B} \| &= \begin{Vmatrix} 0 & i \\ -i & 0 \end{Vmatrix}, & \| S_{23}{}^{A}{}_{B} \| &= \begin{Vmatrix} i & 0 \\ 0 & -i \end{Vmatrix}, \end{aligned}$$

and hence in this coördinate system

$$S_{12} = -iS_{34}, \qquad S_{13} = iS_{24}, \quad \text{and} \quad S_{14} = iS_{23}.$$

However, these equations may be written as

$$(1.25) \qquad \breve{S}^{\mu\nu} \equiv \frac{1}{2}\frac{1}{\sqrt{g}}\epsilon^{\mu\nu\sigma\tau} S_{\sigma\tau} = S^{\mu\nu},$$

where $g = |g_{\sigma\tau}|$ ($= -1$ in a preferred coördinate system) and $\epsilon^{\mu\nu\sigma\tau}$ is the anti-symmetric tensor which is equal to zero unless all indices are different and then it is equal to $+1$ or -1 if $\mu\nu\sigma\tau$ is an even or odd permutation of 1 2 3 4. Since equations (1.25) are spinor equations, they hold in all coördinate systems.

From equations (1.24) and (1.4) the following equations may be proved by verifying them in the special coördinate system,

$$(1.26) \qquad S^{\mu\nu A}{}_{B} S_{\sigma\tau}{}^{B}{}_{A} = -2\eta^{\mu\nu}_{\sigma\tau},$$

$$S^{\mu\nu A}{}_{B}\bar{g}_{\rho}{}^{\dot{B}}{}_{C} = -\eta^{\mu\nu}_{\rho\tau}\bar{g}^{\tau\dot{A}}{}_{C}, \tag{1.27}$$

and

$$g_{\rho}{}^{\dot{A}}{}_{B}S^{\mu\nu B}{}_{C} = \eta^{\mu\nu}_{\rho\tau}g^{\tau\dot{A}}{}_{C}, \tag{1.28}$$

where

$$\eta^{\mu\nu}_{\sigma\tau} = \delta^{\mu\nu}_{\sigma\tau} + \frac{1}{\sqrt{g}}\,\epsilon^{\mu\nu\xi\eta}g_{\xi\sigma}g_{\eta\tau} \tag{1.29}$$

and $\delta^{\mu\nu}_{\sigma\tau}$ is a generalized Kronecker delta, $\delta^{\mu\nu}_{\sigma\tau} = \delta^{\mu}_{\sigma}\delta^{\nu}_{\tau} - \delta^{\mu}_{\tau}\delta^{\nu}_{\sigma}$.

From equation (1.25) it is evident that the tensor defined by the equations

$$z_{\mu\nu} = S_{\mu\nu AB}F^{AB}, \tag{1.30}$$

where F^{AB} is a symmetric spinor is self dual, that is, $\check{z}_{\mu\nu} = z_{\mu\nu}$. Multiplying equation (1.30) by $S^{\mu\nu CD}$ and summing, we obtain as a consequence of equations (1.20)

$$F^{CD} = \tfrac{1}{8}z_{\mu\nu}S^{\mu\nu CD}. \tag{1.31}$$

Equations (1.30) and (1.31) give explicitly the (1-1) correspondence between symmetric spinors and self dual tensors. Since equations (1.31) are inverse to (1.30) the latter must be identically satisfied when the former are substituted into them. Using this fact we obtain equations (1.26). This gives an alternate proof of that identity.

Consider the determinant of the symmetric spinor F_{AB}, we have

$$| F_{AB} | \equiv \tfrac{1}{2}F^{AB}F_{AB} = \tfrac{1}{16}z_{\sigma\tau}z^{\sigma\tau}. \tag{1.32}$$

Hence F_{AB} is singular if and only if $z_{\sigma\tau}z^{\sigma\tau} = 0$. Since any symmetric spinor may be written as

$$F_{AB} = \tfrac{1}{2}(\psi_A\varphi_B + \psi_B\varphi_A) \tag{1.33}$$

we have that $\psi_A = \rho\varphi_A$ if and only if $z_{\sigma\tau}z^{\sigma\tau} = 0$. Hence a self dual tensor determines a simple spinor if and only if $z_{\sigma\tau}z^{\sigma\tau} = 0$.

We conclude this section by noting that an arbitrary real antisymmetric tensor $f_{\sigma\tau}$ determines a self dual tensor by the relations

$$z_{\mu\nu} = f_{\mu\nu} + \check{f}_{\mu\nu} = \tfrac{1}{2}(z_{\mu\nu} + \bar{z}_{\mu\nu}) + \tfrac{1}{2}(z_{\mu\nu} - \bar{z}_{\mu\nu}). \tag{1.34}$$

Since $\check{f}_{\mu\nu}$ is imaginary, if $f_{\mu\nu}$ is real we see that a self dual antisymmetric tensor always determines a real antisymmetric tensor by the relations

$$f_{\mu\nu} = \tfrac{1}{2}(z_{\mu\nu} + \bar{z}_{\mu\nu}). \tag{1.35}$$

It may be proved that

$$\tfrac{1}{2}z^{\mu\nu}\bar{z}_{\nu\tau} = f^{\mu\nu}f_{\nu\tau} + \tfrac{1}{4}f^{\sigma\rho}f_{\sigma\rho}\delta^{\mu}_{\tau}. \tag{1.36}$$

If $f_{\mu\nu}$ is the antisymmetric tensor corresponding to an electromagnetic field, the right hand side of (1.36) is the expression for the stress energy tensor of this field.

2. Tensors Determined by the Solutions of the Dirac Equations

In the preceding section we considered coördinate transformations in the spin space and those in the space S_4 as unrelated. From this viewpoint the invariance in form of the Dirac equations under arbitrary coördinate transformations and constant spin transformations follows immediately. (The restriction to constant spin transformations may be removed by replacing the ordinary derivative by a covariant derivative: notes Chapter III.) The Dirac equations, however, have another invariance property, namely, the coefficients $g^{\sigma\dot{A}}{}_B$ are numerically invariant under an arbitrary Lorentz transformation if we replace ψ^A and φ^A by suitable linear combinations. This type of transformation and the corresponding Lorentz transformation will be called geometric transformations and may be distinguished from coördinate transformations by considering them as a permutation of the points of the space in contrast to a renaming of their coördinates.

This property of the Dirac equations is due to the isomorphism between the antiprojective group in the spin-space and the Lorentz group in S_4 which may be stated as follows:[8]

For every proper Lorentz transformation with coefficients $L^\sigma{}_\tau$ in S_4 there are exactly two unimodular matrices $P = \| P^A{}_B \|$ and $-P$ which satisfy

$$g^\sigma = \overline{P} g^\tau P^{-1} L^\sigma{}_\tau, \tag{2.1}$$

and for every improper Lorentz transformation there are exactly two unimodular matrices $P = \| P^{\dot{A}}{}_B \|$ and $-P$ which satisfy

$$g^\sigma = P \bar{g}^\tau \overline{P}^{-1} L^\sigma{}_\tau. \tag{2.2}$$

In both cases the matrices P determine the Lorentz transformation uniquely and give a double valued representation of the Lorentz group. It should be noticed that the matrices of the type $P = \| P^A{}_B \|$ determine a linear transformation of the form

$$\varphi^* = P\varphi,$$

whereas the matrices of the type $P = \| P^{\dot{A}}{}_B \|$ determine an antilinear transformation of the type

$$\bar{\varphi}^* = P\varphi.$$

The numerical invariance of the coefficients of the Dirac equations under Lorentz transformations is readily proved by means of equations (2.1) and (2.2). Thus if we perform the proper Lorentz transformation

$$x^{\sigma *} = L^\sigma{}_\tau x^\tau,$$

---

[8] Notes page 19.

equations (1.5) become

$$g^{\sigma}L^{\tau}{}_{\sigma}p^{*}_{\tau}\psi = -imc\bar{\varphi},$$

$$\bar{g}^{\sigma}L^{\tau}{}_{\sigma}p^{*}_{\tau}\bar{\varphi} = -imc\psi,$$

where

$$p^{*}_{\sigma} = \frac{h}{i}\left(\frac{\partial}{\partial x^{\sigma *}} - \frac{e}{c}\,e^{\tau}{}_{\sigma}\varphi_{\tau}\right) \quad \text{and} \quad e^{\tau}{}_{\sigma}L^{\sigma}{}_{\rho} = \delta^{\tau}{}_{\rho}.$$

If we now make the spin transformation

(2.3) $$\psi^{*} = P\psi, \qquad \varphi^{*} = P\varphi,$$

these equations become (1.5) in the new quantities. However, if $L^{\tau}{}_{\sigma}$ is an improper Lorentz transformation, we must make the substitution

(2.4) $$\psi^{*} = \bar{P}\bar{\varphi}, \qquad \varphi^{*} = \bar{P}\bar{\psi},$$

to regain equations (1.5).

We now examine the following quantities determined by two simple spinors ψ_A and φ^A:

(2.5) $$T = \psi^{A}\varphi_{A},$$

(2.6) $$A^{\rho} = \bar{\psi}^{\dot{A}}g^{\rho}{}_{\dot{A}B}\psi^{B}, \qquad B^{\rho} = \bar{\varphi}^{\dot{A}}g^{\rho}{}_{\dot{A}B}\varphi^{B}, \qquad C^{\rho} = \bar{\psi}^{\dot{A}}g^{\rho}{}_{\dot{A}B}\varphi^{B},$$

and

(2.7) $$P^{\sigma\tau} = S^{\sigma\tau}{}_{AB}\psi^{A}\varphi^{B}, \qquad M^{\sigma\tau} = S^{\sigma\tau}{}_{AB}\psi^{A}\psi^{B}, \qquad R^{\sigma\tau} = S^{\sigma\tau}{}_{AB}\varphi^{A}\varphi^{B}.$$

Under transformations of spin coördinates these are all scalars. Under coördinate transformations in S_4 they transform as tensors of the type indicated by their indices (T is a scalar, A^{ρ} a vector, etc.). From equation (1.13) it is evident that A_{ρ} and B_{ρ} are real; the remaining are complex, and $P^{\sigma\tau}$, $M^{\sigma\tau}$, and $R^{\sigma\tau}$ are self dual antisymmetric tensors as follows from (1.25). A_{ρ}, B_{ρ}, C_{ρ}, $M_{\sigma\tau}$, and $R_{\sigma\tau}$ have been treated by Whittaker and Ruse has discussed $P_{\sigma\tau}$.

Consider the behavior of these quantities under geometric transformations in the spin space: A transformation of the type (2.3) induces on these seven quantities the tensor transformation indicated by the indices. For example

$$A^{\rho *} = \bar{\psi}^{\dot{A}*}g^{\rho}{}_{\dot{A}B}\psi^{B*} = \bar{P}^{\dot{A}}{}_{\dot{C}}\bar{\psi}^{\dot{C}}g^{\rho}{}_{\dot{A}B}P^{B}{}_{D}\psi^{D} = L^{\rho}{}_{\sigma}\bar{\psi}^{\dot{A}}g^{\sigma}{}_{\dot{A}B}\psi^{B} = L^{\rho}{}_{\sigma}A^{\sigma},$$

where $L^{\sigma}{}_{\tau}$ is the proper Lorentz transformation determined by equations (2.1); since the fact that P is a unimodular matrix implies $\epsilon_{AB}P^{A}{}_{C}P^{B}{}_{D} = \epsilon_{CD}$ whence $P_{AB} = -P^{-1}_{BA}$. Similarly it can be shown that B_{ρ} and C_{ρ} transform as vectors, $P_{\sigma\tau}$, $M_{\sigma\tau}$ and $R_{\sigma\tau}$ transform as antisymmetric tensors and T is a scalar.

However, a transformation of the type (2.4) has a different effect on these quantities. Thus the antilinear transformation:

$$\psi^{A} \to \psi^{A*} = \bar{P}^{\dot{A}}{}_{\dot{B}}\bar{\varphi}^{\dot{B}}, \qquad \varphi^{A} \to \varphi^{A*} = \bar{P}^{\dot{A}}{}_{\dot{B}}\bar{\psi}^{\dot{B}}$$

sends

$$
\begin{aligned}
T \rightarrow T^* &= \psi^{A*} \varphi_A^* = -\bar{P}^{\dot{A}}{}_{B}\bar{\varphi}^B \bar{P}_{\dot{A}}{}^{C}\bar{\psi}_C = \bar{\varphi}^B \bar{\psi}_B = -\bar{T}, \\
A^\rho \rightarrow A^{\rho*} &= L^\rho{}_\sigma B^\sigma, \qquad P^{\sigma\tau} \rightarrow P^{\sigma\tau*} = L^\sigma{}_\mu L^\tau{}_\nu \bar{P}^{\mu\nu}, \\
B^\rho \rightarrow B^{\rho*} &= L^\rho{}_\sigma A^\sigma, \qquad M^{\sigma\tau} \rightarrow M^{\sigma\tau*} = L^\sigma{}_\mu L^\tau{}_\nu \bar{R}^{\mu\nu}, \qquad (2.8) \\
C^\rho \rightarrow C^{\rho*} &= L^\rho{}_\sigma \bar{C}^\sigma, \qquad R^{\sigma\tau} \rightarrow R^{\sigma\tau*} = L^\sigma{}_\mu L^\tau{}_\nu \bar{M}^{\mu\nu},
\end{aligned}
$$

where $L^\sigma{}_\tau$ is now the improper Lorentz transformation determined by equation (2.2).

From these relations it is evident that the quantities

$$\Delta = T - \bar{T}, \qquad J^\sigma = A^\sigma + B^\sigma, \qquad (2.9)$$

and

$$f_{\sigma\tau} = P_{\sigma\tau} + \bar{P}_{\sigma\tau}$$

undergo a tensor transformation of the type indicated by their indices whenever ψ and φ undergo a transformation of the type (2.3) or (2.4). It is possible to form quantities which have an induced tensor transformation law out of combinations of the quantities C^ρ, $M^{\sigma\tau}$, and $R^{\sigma\tau}$ and their conjugates. However, these are not gauge invariant quantities if ψ_A and φ_A are solutions of the Dirac equation, whereas those given by equation (2.9) are. This follows from the fact that if φ_σ is replaced by $\varphi_\sigma + \partial S/\partial x^\sigma$, the Dirac equations (1.1) are unaltered if we replace ψ^A by $e^{(ie/ch)S}\psi^A$ and φ^A by $e^{-(ie/ch)S}\varphi^A$. Because C_ρ, $M_{\sigma\tau}$, and $R_{\sigma\tau}$ and their conjugates are not gauge invariant, neither they themselves nor quantities built from them by linear combinations can have a direct physical interpretation. However, gauge invariant quadratic combinations can be formed and these may have a physical interpretation.

The seven quantities defined by equations (2.5), (2.6) and (2.7) are not all independent but satisfy identities some of which have been given by Ruse and Whittaker obtained from equations (1.16), (1.22) and (1.23). These are the same as the quadratic identities given by Pauli[9] for the tensors built from four component spinors. In addition Whittaker's "Catalytic" relations follow from the latter equations.

3. Tensor Equations Equivalent to the Dirac Equations

The tensor equations shown by Whittaker to be equivalent to the Dirac equations are

$$
\begin{aligned}
\Omega^\tau = \frac{1}{2}(M^{\tau\sigma} + R^{\tau\sigma})_{,\sigma} - \frac{ie}{hc}(M^{\tau\sigma} - R^{\tau\sigma})\varphi_\sigma \qquad (3.1) \\
- (\psi_C \psi^C_{,\sigma} + \varphi_C \varphi^C_{,\sigma})g^{\sigma\tau} + \frac{mc}{h}(C^\tau + \bar{C}^\tau) = 0.
\end{aligned}
$$

[9] W. Pauli: "Beitrage zur Mathematischen Theorie der Diracschen Matrizen," *Zeeman Verhandelingen*, Nijhoff (1935).

The quantity Ω^r transforms as a contravariant vector under arbitrary coördinate transformations and as a scalar under constant spin coördinate transformations. A constant geometric transformation of the type (2.3) in the spin spaces sends

$$\Omega \rightarrow \Omega^{r*} = L^r{}_s\Omega^s$$

and, one of the tvpe (2.4) sends

$$\Omega^r \rightarrow \Omega^{r*} = L^r{}_s\bar{\Omega}^s$$

provided we perform the corresponding Lorentz transformation in S_4. Hence equation (3.4) does not as it stands have an induced vector transformation law. However, it is equivalent to a pair of equations which do, since $\Omega^r = 0$ is equivalent to the two statements

$$R^r = \tfrac{1}{2}(\Omega^r + \bar{\Omega}^r) = 0,$$

and

$$I^r = \frac{1}{2i}(\Omega^r - \bar{\Omega}^r) = 0.$$

R^r has the induced transformation law of a contravariant vector and I^r has the induced transformation law of a pseudo vector, that is, of the dual third order antisymmetric tensor associated with every vector.

As was noted before, the quantities which enter into equations (3.1) are not gauge invariant and therefore cannot have a physical interpretation. Ruse obtained another set of equations, which do have a physical interpretation and which are equivalent to the Dirac equations, they are the real and imaginary parts of Λ^r set equal to zero where

$$\Lambda^r = -\frac{h}{i}P^{r\sigma}_{.\,,\sigma} + g^{r\sigma}(\varphi_B p_\sigma \psi^B + \psi^C \bar{p}_\sigma \varphi_C) + imcJ^r. \tag{3.2}$$

The transformation properties of Λ^r are the same as those of Ω^r.

4. Maxwell's Equations in Spinor Form

Laporte and Uhlenbeck[3] were the first to use the (1-1) correspondence between self dual tensors and symmetric spinors to obtain the spinor form of the Maxwell equations. The formalism of the first section enables us to obtain these results quite readily. We give them here so as to compare them with the equations recently proposed by Dirac[4] and used by Kemmer[5] for material particles with spin one.

The Maxwell equations in free space may be written as

$$z^{\sigma\tau}{}_{,\tau} = I^\sigma, \tag{4.1}$$

where $z^{\sigma\tau}$ is a self dual complex antisymmetric tensor and I^σ is a timelike real vector, the current vector. In terms of the four potential vector φ_σ we have

$$z_{\sigma\tau} = \tfrac{1}{2}\eta^{\mu\nu}_{\sigma\tau}(\varphi_{\mu,\nu} - \varphi_{\nu,\mu}). \tag{4.2}$$

If we write

$$z_{\sigma\tau} = F_{AB}S_{\sigma\tau}{}^{AB} \quad \text{and} \quad \varphi_\sigma = g_\sigma{}^{\dot{A}B}\Phi_{\dot{A}B}\,,$$

equations (4.2) become on multiplying by $S^{\sigma\tau}{}_{CD}$

$$8F_{CD} = 2S^{\mu\nu}{}_{CD}((g_\mu{}^{\dot{A}B}\Phi_{\dot{A}B})_{,\nu} - (g_\nu{}^{\dot{A}B}\Phi_{\dot{A}B})_{,\mu}),$$

which may be written as

$$F_{CD} = -\tfrac{1}{2}(g^{\nu\dot{A}}{}_D\Phi_{\dot{A}C,\nu} + g^{\nu\dot{A}}{}_C\Phi_{\dot{A}D,\nu}), \tag{4.3}$$

where we have used equation (1.22).

Equations (4.1) may be written as

$$S^{\sigma\tau AB}F_{AB,\tau} = g^{\sigma\dot{A}B}I_{\dot{A}B}\,.$$

Multiplying this by $g_{\sigma\dot{C}D}$ and using (1.22), we obtain

$$g^{\tau\dot{C}}{}_B F^{BD}{}_{,\tau} = I^{\dot{C}D}. \tag{4.4}$$

Equation (4.4) is fully equivalent to equation (4.1) since the latter may be derived from the former by using the formulas from the latter part of the first section.

The equations proposed by Dirac for a free particle of spin 1 are in the notation used here

$$\begin{gathered} g^{\sigma\dot{A}B}A_{BC,\sigma} = KB^{\dot{A}}{}_C\,, \\ g^\sigma{}_{\dot{A}C}B^{\dot{A}}{}_{D,\sigma} = -KA_{CD}\,, \\ K = \frac{mc}{h}, \end{gathered} \tag{4.5}$$

where m = mass of the particle and A_{BC} is a symmetric spinor. It is readily verified that as a consequence of equations (4.5) both A_{BC} and $B^{\dot{A}}{}_C$ satisfy second order wave equations. We see that aside from constant factors equations (4.5) are exactly the same as (4.3) and (4.4) provided we identify $I_{\dot{A}B}$ and $\Phi_{\dot{A}B}$ in the latter two.

We will now determine the tensor equations equivalent to (4.5). Since A_{BC} is symmetric, we may write

$$A_{BC} = \tfrac{1}{4}A_{\mu\nu}S^{\mu\nu}{}_{BC} \quad \text{and} \quad B^{\dot{A}}{}_C = \varphi_\sigma g^{\sigma\dot{A}}{}_C \tag{4.6}$$

where $A_{\mu\nu}$ is a self dual antisymmetric tensor and φ_σ is a covariant vector, which is real if $B_{\dot{A}C}$ is Hermitian. The first of equations (4.5) becomes

$$g^{\sigma\dot{A}B}(\tfrac{1}{4}A_{\mu\nu}S^{\mu\nu}{}_{BC})_{,\sigma} = \tfrac{1}{4}g^{\sigma\rho}A_{\mu\nu,\sigma}\eta^{\mu\nu}_{\rho\tau}g^{\tau\dot{A}}{}_C = g^{\sigma\rho}A_{\rho\tau,\sigma}g^{\tau\dot{A}}{}_C = K\varphi_\tau g^{\tau\dot{A}}{}_C\,,$$

or

$$A^{\sigma\tau}{}_{,\sigma} = K\varphi^\tau. \tag{4.7}$$

The second of equations (4.5) becomes

$$g^{\sigma}{}_{\dot{A}C}\, g^{\tau\dot{A}}{}_{D}\varphi_{\tau,\sigma} = -\frac{K}{4} A_{\mu\nu} S^{\mu\nu}{}_{CD},$$

or

$$S^{\tau\sigma}{}_{CD}\varphi_{\tau,\sigma} + g^{\tau\sigma}\epsilon_{CD}\varphi_{\tau,\sigma} = -\frac{K}{4} A_{\mu\nu} S^{\mu\nu}{}_{CD}. \tag{4.8}$$

Hence we must have

$$g^{\sigma\tau}\varphi_{\tau,\sigma} = 0, \tag{4.9}$$

which follows from (4.7). Multiplying equations (4.8) by $S_{\rho\lambda}{}^{CD}$ and summing, we obtain

$$-KA_{\rho\lambda} = \eta_{\rho\lambda}^{\tau\sigma}\varphi_{\tau,\sigma}\,. \tag{4.10}$$

Equations (4.7) and (4.10) together are equivalent to the set (4.5), since the latter may be derived from these equations in the manner in which we derived the spinor form of the Maxwell equations. Equations (4.7) and (4.10) differ from the equations proposed by Proca[6] in that the antisymmetric self dual tensor $A_{\rho\lambda}$ is not the curl of a vector φ_τ but only the self dual part of the curl of the vector. If the vector were real, the equations (4.7), (4.9) and (4.10) could be decomposed into two sets, which would be exactly the Proca equations for real φ_ν.

The tensor equations equivalent to the spinor equations

$$\begin{aligned} g^{\sigma\dot{A}B}p_\sigma A &= m'B^{\dot{A}B}, \\ g^{\sigma\dot{A}B}p_\sigma B_{\dot{A}B} &= m''A, \end{aligned} \tag{4.11}$$

which were proposed by Kemmer are

$$\begin{aligned} p_\sigma\varphi^\sigma &= m''A, \\ p_\sigma A &= 2m'\varphi_\sigma\,, \end{aligned} \tag{4.12}$$

where $B^{\dot{A}B} = \varphi_\sigma g^{\sigma\dot{A}B}$.

UNIVERSITY OF WASHINGTON,
SEATTLE, WASHINGTON.

ANNALS OF MATHEMATICS
Vol. 40, No. 4, October, 1939

A CORRECTION

R. P. BOAS, JR.

(Received December 1, 1938)

In my paper, Representations for entire functions of exponential type,[1] I made the remark that a function $f(z)$, of exponential type R, and bounded on the real axis, can have the form

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t),$$

with $\alpha(t)$ not of the class Lip 0; but my assertion that "trigonometric polynomials furnish examples" is obviously incorrect, since Lip 0 is naturally to be interpreted as the class of bounded functions, not the class of continuous functions. A correct example is

$$\begin{aligned} f(z) &= \int_{0+}^{\pi/2} \sin zt\, d(\log \sin t) \\ &= \int_{0}^{\pi/2} \sin zt \operatorname{ctn} t\, dt \\ &= -z \int_{0}^{\pi/2} \log \sin t \cos zt\, dt. \end{aligned}$$

That $f(x)$ is bounded for large $|x|$ follows from the second integral, which is the Dirichlet integral for a function of bounded variation; the third integral shows that $f(x)$ is bounded for small $|x|$.

CAMBRIDGE, ENGLAND.

[1] These *Annals*, 39 (1938), pp. 269–286; p. 284.

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 40, NO. 1
JANUARY, 1939

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscription price, five dollars a volume (four numbers) in advance. Single copies one dollar and fifty cents. An arrangement has been made with the Mathematical Association of America by which an individual member of the Association may subscribe to the Annals of Mathematics at one-half the regular price. Subscriptions, orders for back numbers, and changes of address should be sent

from the United States and Canada to Mount Royal and Guilford Avenues, Baltimore, Maryland, or preferably to the Princeton University Press, Princeton, N. J.

from all other countries, to Nordemann Publishing Company, N. V., 243 O. Z. Voorburgwal, Amsterdam C., Holland.

Manuscripts and all editorial correspondence should be addressed to the Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced. Footnotes should be numbered consecutively and typed (double-spaced) on a separate sheet; they will not be inserted in the text until page proofs are made.

Figures and diagrams should be drawn in India ink on separate sheets of paper and twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Authors ordinarily receive galley proofs only. Stylistic changes must be avoided in proof; excessive corrections will be charged to the authors. Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors receive fifty reprints each.

Suggestions to Authors

Footnotes should be reduced to a minimum or replaced whenever possible by a bibliography at the end of the paper; formulae in footnotes should be avoided. Authors are requested to keep in mind the typographical difficulties of complicated formulae. For example, "exp" may be used to avoid complicated exponents and a power to replace the troublesome radical sign. Symbols set one on top of another present a typographical difficulty which in most cases can easily be avoided by a stepwise arrangement.

Princeton University Press. Agent for countries outside the United States and Canada: Nordemann Publishing Co. N. V., O. Z. 243 Voorburgwal, Amsterdam.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in United States of America

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 40, NO. 2

APRIL, 1939

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscription price, five dollars a volume (four numbers) in advance. Single copies one dollar and fifty cents. An arrangement has been made with the Mathematical Association of America by which an individual member of the Association may subscribe to the Annals of Mathematics at one-half the regular price. Subscriptions, orders for back numbers, and changes of address should be sent

from the United States and Canada to Mount Royal and Guilford Avenues, Baltimore, Maryland, or preferably to the Princeton University Press, Princeton, N. J.

from all other countries, to Nordemann Publishing Company, N. V., 243 O. Z. Voorburgwal, Amsterdam C., Holland.

Manuscripts and all editorial correspondence should be addressed to the Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced. Footnotes should be numbered consecutively and typed (double-spaced) on a separate sheet; they will not be inserted in the text until page proofs are made.

Figures and diagrams should be drawn in India ink on separate sheets of paper and twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Authors ordinarily receive galley proofs only. Stylistic changes must be avoided in proof; excessive corrections will be charged to the authors. Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors receive fifty reprints each.

Suggestions to Authors

Footnotes should be reduced to a minimum or replaced whenever possible by a bibliography at the end of the paper; formulae in footnotes should be avoided. Authors are requested to keep in mind the typographical difficulties of complicated formulae. For example, "exp" may be used to avoid complicated exponents and a power to replace the troublesome radical sign. Symbols set one on top of another present a typographical difficulty which in most cases can easily be avoided by a stepwise arrangement.

Princeton University Press. Agent for countries outside the United States and Canada: Nordemann Publishing Co. N. V., O. Z. 243 Voorburgwal, Amsterdam.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in United States of America

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 40, NO. 3

JULY, 1939

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscription price, five dollars a volume (four numbers) in advance. Single copies one dollar and fifty cents. An arrangement has been made with the Mathematical Association of America by which an individual member of the Association may subscribe to the Annals of Mathematics at one-half the regular price. Subscriptions, orders for back numbers, and changes of address should be sent

from the United States and Canada to Mount Royal and Guilford Avenues, Baltimore, Maryland, or preferably to the Princeton University Press, Princeton, N. J.

from all other countries, to Nordemann Publishing Company, N. V., 243 O. Z. Voorburgwal, Amsterdam C., Holland.

Manuscripts and all editorial correspondence should be addressed to the Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced. Footnotes should be numbered consecutively and typed (double-spaced) on a separate sheet; they will not be inserted in the text until page proofs are made.

Figures and diagrams should be drawn in India ink on separate sheets of paper and twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Authors ordinarily receive galley proofs only. Stylistic changes must be avoided in proof; excessive corrections will be charged to the authors. Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors receive fifty reprints each.

Suggestions to Authors

Footnotes should be reduced to a minimum or replaced whenever possible by a bibliography at the end of the paper; formulae in footnotes should be avoided. Authors are requested to keep in mind the typographical difficulties of complicated formulae. For example, "exp" may be used to avoid complicated exponents and a power to replace the troublesome radical sign. Symbols set one on top of another present a typographical difficulty which in most cases can easily be avoided by a stepwise arrangement.

Princeton University Press. Agent for countries outside the United States and Canada: Noordemann Publishing Co. N. V., O. Z. 243 Voorburgwal, Amsterdam.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in United States of America

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 40, NO. 4

OCTOBER, 1939

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscription price, five dollars a volume (four numbers) in advance. Single copies one dollar and fifty cents. An arrangement has been made with the Mathematical Association of America by which an individual member of the Association may subscribe to the Annals of Mathematics at one-half the regular price. Subscriptions, orders for back numbers, and changes of address should be sent

from the United States and Canada to Mount Royal and Guilford Avenues, Baltimore, Maryland, or preferably to the Princeton University Press, Princeton, N. J.

from all other countries, to Nordemann Publishing Company, N. V., 243 O. Z. Voorburgwal, Amsterdam C., Holland.

Manuscripts and all editorial correspondence should be addressed to the Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced. Footnotes should be numbered consecutively and typed (double-spaced) on a separate sheet; they will not be inserted in the text until page proofs are made.

Figures and diagrams should be drawn in India ink on separate sheets of paper and twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Authors ordinarily receive galley proofs only. Stylistic changes must be avoided in proof; excessive corrections will be charged to the authors. Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors receive fifty reprints each.

SUGGESTIONS TO AUTHORS

Footnotes should be reduced to a minimum or replaced whenever possible by a bibliography at the end of the paper; formulae in footnotes should be avoided. Authors are requested to keep in mind the typographical difficulties of complicated formulae. For example, "exp" may be used to avoid complicated exponents and a power to replace the troublesome radical sign. Symbols set one on top of another present a typographical difficulty which in most cases can easily be avoided by a stepwise arrangement.

Princeton University Press. Agent for countries outside the United States and Canada: Noordemann Publishing Co. N. V., O. Z. 243 Voorburgwal, Amsterdam.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in United States of America

TABLE OF CONTENTS